Secretin interactions in the type II secretion system

Gu, Shuang

Queen Mary University of London
School of Biological and Chemical Sciences

Secretin interactions in the type II secretion system

A thesis submitted by

Shuang Gu

In partial fulfillment of the requirement of
UNIVERSITY OF LONDON
For the degree of Doctor of Philosophy

August 2011

Declaration

This thesis is a presentation of my original research work. Wherever contributions of others are involved, every effort is made to indicate this clearly, with due reference to the literature, and acknowledgement of collaborative research and discussions.

Acknowledgments

I would like to begin these acknowledgements by sincerely thanking my supervisor Professor R W Pickersgill for giving me this amazing opportunity to work in his lab and having such faith in me. He gives me all the freedom to explore, which afforded me the independence freely to pursue my interests. Nevertheless he has always been on hand to help me out and keep me on the track. I thank Richard for his open-mindedness for supporting me to complete this NMR based thesis in his X-ray crystallography lab.

This thesis would not be possible without expertise assistance and support from MRC NMR centre (NIMR), especially Dr Geoff Kelly. Yours unrelenting enthusiasm, patience and good will has inspired me no end.

It was lucky for me to land in a very lively and helpful lab. We shared lots of thoughts and laughs. Thank you all in the lab, Tom, Arefeh, Teng, Yumi, Kerem, Allan, Saima, Becky, Nadine and Claire.

Last but not least, a huge thank you goes to my friends and family, especially my Mum and Dad. Although we are far apart, your persistent encouragement and support is with me all the time. I hope I've done you all proud.

The experience of undertaking this PhD will be with me forever.

Abstract

The type II secretion system (T2SS) is the major terminal branch of the general secretory pathway. It is composed of 12-15 proteins, most in multiple copies, and spans the inner and outer membranes of Gram-negative bacteria. The T2SS secretin subunits form a large dodecameric torus-like structure in the outer membrane. The secretin is the only essential component in the outer membrane and secreted proteins and virulence factors pass through the pore in the toroidal secretin dodecamer and out into the environment.

The interaction between the secretin and its partners plays a key role in regulation of the T2SS. The interaction between the so-called homology region of the inner-membrane protein GspC (GspC-HR) and secretin provides the structural and functional integrity of the secretion machinery across the two cell membranes. The interaction between secretin and its pilotin translocates the secretin subunits to the outer membrane.

In this Thesis, the interactions between secretin and its partners are studied at molecular level. The GspC-HR structure is solved using NMR spectroscopy. Its interaction with secretin (GspD) is elucidated using several biochemical and biophysical approaches and a model of the complex is proposed. Also, the interaction between secretin (GspD) and pilotin (GspS) is further charicterisied. An 18 residues secretin sequence is identified as responsible for interacting with pilotin. Upon binding to the pilotin, the unstructured secretin forms a helical structure.

TABLE OF CONTENT

DECLARATION **2**

ACKNOWLEDGMENTS **3**

ABSTRACT **4**

TABLE OF CONTENT **5**

LIST OF FIGURES **9**

LIST OF TABLES **11**

ABBREVIATIONS **12**

CHAPTER 1 INTRODUCTION **14**

1.1 Bacteria secretion systems **14**

1.1.2 Architecture of the Gram-negative bacteria cell envelope 15

1.1.3 Secretion in Gram-negative bacteria 16

1.1.4 Type II secretion pathogens 17

1.1.5 Type II secretion related disease 18

1.2 Type II secretion system components **19**

1.2.1 Structure of T2SS components 21

1.3.1 Structure of secretins 24

1.3.2 T2SS secretin 27

1.4 Type II secretin and its partners **30**

1.4.1 GspC and GspD; gatekeepers of the T2SS 30

1.4.2 GspC is a key player in T2SS 31

1.5 Secretin's pilotin **33**

1.5.1 Pilotin lipidation 34

1.5.2 Lol system 36

1.5.3 Structure of pilotin 38

1.6 Aims of this study **38**

CHAPTER 2 MATERIAL AND METHODS **40**

2.1 Molecular biology **40**

2.1.1 Plasmid construction 40

2.1.2 Bacterial strains 40
2.1.3 DNA manipulation and analysis 40
2.1.4 Polymerase chain reaction (PCR) 41
2.1.5 Restriction enzyme digestion 42
2.1.6 Ligation reactions 43
2.1.7 Transformation 43
2.1.8 Protein expression 43
2.2 Protein purification techniques 44
2.2.1 Bacterial lysis 44
2.2.2 His tag purification 44
2.2.3 GST tag purification 45
2.2.4 Purifying protein from the periplasm 46
2.2.5 Size exclusion chromatography 46
2.2.6 SDS-PAGE 47
2.2.7 NMR sample preparation 48
2.2.8 Protein concentration determination 48
2.3 Biochemical and biophysical techniques 49
2.3.2 Mass spectrometry 49
2.3.3 N-terminal sequencing 50
2.3.4 Dynamic light scattering 50
2.3.5 Circular dichroism 51
2.3.6 Fluorescence spectroscopy 52
2.3.7 Preparation of OutC and OutD coexpression pull down assays 53
2.3.8 Thermofluor 53
CHAPTER 3 NMR SPECTROSCOPY: MATERIALS AND METHODS 55
3.1 Overview 55
3. 2 NMR spectroscopy 55
3.2.1 Spectroscopy 55
3.2.2 Spectral analysis and resonance assignment 57
3.3 Dynamic studies by NMR 58
3.3.1 Longitudinal and transverse relaxation 59
3.3.2 The Nuclear Overhauser Effect (NOE) 60
3.4 Secondary structure prediction by TALOS 61
3.5 Residual dipolar coupling 62

3.6 ARIA 1.2 ..... **63**
3.7 NMR titration **64**
3.8 HADDOCK **64**
CHAPTER 4: SOLUTION STRUCTURE OF OUTC-HRF3 ..... **65**
4.1 Overview **65**
4.2 OutC- HRF3 structure identification **65**
4.3 OutC-HRF3 expression **69**
4.3.1 Sub-cloning outC-hrf3 69
4.3.2 OutC-HRF3 expression 70
4.3.3 OutC-HRF3 purification 70
4.4 Dynamic light scattering **71**
4.5 OutC-HRF3 solution structure **73**
4.5.1 Determining the architecture of OutC-HRF3 using NMR spectroscopy 73
4.5.3 OutC-HRF3 secondary structure prediction 76
4.5.6 CS-ROSETTA structure prediction 80
4.5.7 Hydrogen bond restraints 82
4.5.8 Residual dipolar coupling 84
4.5.9 Resonance assignment of OutC-HRF3 86
4.5.10 NMR structure calculation 86
4.5.11 OutC-HRF3 solution structure evaluation 87
4.6 Summary and discussion **94**
CHAPTER 5 STUDY OF THE INTERACTION BETWEEN OUTC-HRF3 AND OUTD-N0 ..... **95**
5.1 Overview **95**
5.2 Production of OutD domains **95**
5.2.1 Cloning of pET-14b outD-n0 96
5.3 OutC-HRF3 interacts with OutD-N0 ..... **99**
5.3.1 OutC-HRF3 does not interact N1 or N2 of OutD 101
5.4 NMR study of OutD-N0 ..... **103**
5.4.1 Backbone assignment of OutD-N0 103
5.4.2 OutD-N0 secondary structure prediction 103
5.4.4 OutD-N0 structure modulation 107
5.4.5 GspD N0 and N1 domain arrangement 107

5.6 Model of OutC-HRF3 and OutD-N0 interaction by HADDOCK 119
5.7 Interaction study of OutC-HRF3 and OutD-N0 by thermofluor 122
5.8 OutC-HRF3 and OutD-N0 interaction by His Trap-pull down assay 123
5.9 Summary and discussion 126
CHAPTER 6 INTERACTION BETWEEN OUTS AND OUTD 128
6.1 Overview 128
6.2 Bioinformatic studies of OutS and OutD-Cter 128
6.3 Expression and purification of OutS 131
6.4 Expression and purification of OutD-Cter 132
6.4.1 Purification of OutD-Cter* with GST tag 133
6.4.2 Purification of OutD-Cter with His tag and GST Tag 133
6.5 Biophysical studies of OutS and OutD-Cter 135
6.5.1 OutS DLS 136
6.5.2 OutS CD 136
6.5.3 OutD-Cter CD 137
6.5.4 Interaction studies of OutS and OutD-Cter by CD 138
6.6 NMR study of OutS and OutD-Cter 140
6.6.1 OutS 2D HSQC spectrum 140
6.6.2 OutD-Cter 2D HSQC spectrum 141
6.7 Interaction between OutS and OutD-Cter* 142
6.8 OutD$^{692-708}$ interacting with OutS 146
6.9 Summary and discussion 152
CHAPTER 7 OVERALL CONCLUSION 155
REFERENCES 157
APPENDICIES 165

List of Figures

Figure1.1 Schematic representation of bacterium cell envelope **14**
Figure 1.2 Gram-negative bacteria cell envelope structure **15**
Figure 1.3 Secretion systems used by Gram-negative bacteria **17**
Figure 1.4 Model of the T2SS and known high-resolution component structures **22**
Figure 1.5 High-resolution model of the T2SS **23**
Figure 1.6 Secretins in Gram-negative bacteria **25**
Figure 1.7 The secretin architecture is conserved in different secretion systems **26**
Figure 1.8 GspD modular organization and structure **28**
Figure 1.9 Schematic diagram showing the role of GspC **32**
Figure 1.10 Biogenesis of lipoproteins **35**
Figure 1.11 Sorting and outer membrane localization of lipoproteins by the Lol system. **37**
Figure 1.12 Pilotins in Gram-negative bacteria **38**

Figure 2.1 Typical far-UV CD spectra **51**

Figure 3.1 Resonance assignment using HNCACB and HN(CA)CO spectra **58**
Figure 3.2 Schematic representation of T1, T2 and NOE relaxation times as a function of correlation time **60**
Figure 3.3 Protein dihedral angels **62**
Figure 3. 4 Schematic representation of the calculations by the TALOS+ program **62**

Figure 4.1 Limited proteolysis of OutC-HR **67**
Figure 4.2 Sequencing results reveal the sequence of the putative domains **68**
Figure 4.3 Cloning, expression and purification of OutC-HRF3 ..... **69**
Figure 4.4 Recombinantly produced OutC-HRF2 and OutC-HRF3 compared with proteolysis fragments **72**
Figure 4.5 OutC-HRF3 DLS summary **72**
Figure 4.6 Comparison of OutC-HRF3 and OutC-Peri spectra **74**
Figure 4. 7 ^{1}H-^{15}N HSQC spectrum of OutC-HRF3 ..... **75**
Figure 4.8 Secondary structure prediction of OutC-HRF3 ..... **78**
Figure 4.9 T1,T2 and NOE relaxation value of OutC-HRF3 ..... **79**
Figure 4.10 CS-ROSETTA structure energy plot **81**
Figure 4.11 Two views of CS-ROSETTA predicted OutC-HRF3 structure **83**
Figure 4.12 Comparison of OutC-HRF3 and PilP structure. **83**
Figure 4.13 OutC-HRF3 hydrogen bond measurement. **84**
Figure 4.14 Correlations between calculated OutC-HRF3 NMR structure and experimental measured RDC **85**
Figure 4.15 Structure of OutC-HRF3. ..... **90**
Figure 4.16 Sausage representation of best 20 calculated OutC-HRF3 structures **91**

Figure 5.1 Schematic diagram of OutD and its truncated derivatives used in this study..... 96
Figure 5.2 Construction of the pET14b-outD-n0. 96
Figure 5.3 Expression and purification of OutD N-terminal proteins..... 98
Figure 5.4 OutC-hpa titraion with OutD-N0 100
Figure 5.5 The intensity plot of OutC-hpa amide groups during the OutD-N0 titration 101
Figure 5.6 ^{15}N OutC-HRF3 titration with OutD-N1N2..... 102
Figure 5.7 ^{1}H-^{15}N HSQC spectrum of OutD-N0 104
Figure 5.8 Homology based OutD-N0 structure prediction..... 105
Figure 5.9 OutD-N0 T1,T2 and NOE relaxation..... 106
Figure 5.10 Homology based OutD-N0 structure predicttion. 107
Figure 5.11 Interaction between Peri-GspD and the nanobody 108
Figure 5.12 ^{1}H- ^{15}N HSQC of OutD-N0 in the absence and the presence of OutD-N1N2. 109
Figure 5.13 Overlay of ^{1}H-^{15}N HSQC spectrum of OutD-N0 and OutD-N0N1 spectrum 111
Figure 5.15 ^{15}N HSQC of OutC-HRF3 in the absence and the presence of OutD-N0..... 113
Figure 5.16 ^{1}H- ^{15}N HSQC of OutD-N0 in the absence and the presence of OutC-HRF3 114
Figure 5.17 Effect of OutD-N1N2 on the OutC-HRF3 and OutD-N0 mixture 117
Figure 5.19 Interaction surface identified by chemical shift perturbation 120
Figure 5.21 Thermofluor plots..... 123
Figure 5.22 OutC$^{72-162}$ OutD-N0 Ni column pull down assay..... 125

Figure 6.1 Schematic diagrams of OutD C-terminal derivatives used in this Chapter..... 130
Figure 6.2 Family sequence alignment of GspS 130
Figure 6.3 Purification of OutS 132
Figure 6.4Purification of OutD-Cter*..... 134
Figure 6.5 Purification of OutD-Cter..... 135
Figure 6.6 OutS DLS summary. 136
Figure 6.7 Far-UV CD spectrum of OutS..... 137
Figure 6.8 Far-UV CD spectra of OutD-Cter..... 138
Figure 6.9 Far-UV CD spectra of OutS and OutD-Cter mixture..... 140
Figure 6.10 ^{1}H-^{15}N HSQC spectra of OutS..... 141
Figure 6.11 ^{1}H-^{15}N HSQC spectra of OutD-Cter* and backbone assighment 143
Figure 6.12 ^{1}H-^{15}N HSQC spectrum of OutS in the absence or presence of of OutD-Cter 145
Figure 6.13OutD-Cter* and OutD-Cter-cleaved titration with OutS..... 1
Figure 6.14 Far-UV CD spectra of OutD$^{692-708}$. 1
Figure 6 15 OutS titration with OutD$^{692-708}$. 149
Figure 6.16 OutS and OutD$^{691-708}$ Kd determinnation by flourescence spectroscopy 150
Figure 6.17 Proposed OutS OutD binding model..... 151
Figure 6.18A Model of tranlocation of GspS and GspD to the outer membrane via Lol pathway. 153

List of Tables

Table 2.1 Typical PCR reaction mixture **41**
Table 2.2 Typical thermocycler program for PCR reaction. **41**
Table 2.3 Typical restriction enzyme digest for PCR products and vectors **42**
Table 2.4 Typical ligation reaction **43**
Table 2.5 Recipe for SDS-PAGE **47**
Table 2.6 Conditions used in 96 well thermofluor buffer screen. **54**

Abbreviations

Å	Ångstrom
µl/µM/µS	micro-liter/micro-molar/micro-second
τ_c	correlation time
ARIA	ambiguous restraints for iterative assignment
ATP	adenosine-5'-triphosphate
CD	circular dichroism
CS-Rosetta	chemical shift-Rosetta
CV	column volumes
Da/kDa	dalton/kilo dalton
DLS	dynamic light scattering
DNA	deoxyribonucleic acid
DTT	dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
ETEC	enterotoxigenic *Escherichia coli*
Gsp	general secretory pathway
GST	glutathione *S*-transferase
HSQC	heternuclear single quantum coherence
IPTG	Isopropyl β-D-1-thiogalactopyranoside
Kd	dissociation constant
LB	Luria broth
min.	minute
ml/mM/mS	milli-lter/milli-molar/milli-second
nl/nM/nS	nano-liter/nano-molar/nano-second
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
NOESY	nuclear Overhauser effect spectroscopy
PBS	phosphate buffered saline
PCR	polymerase chain reaction
pH	the power of hydrogen ($-\log_{10}$ of the hydrogen ion concentration)
RCI	random coil index
RMSD	root-mean-square deviation
RNA	ribonucleic acid

S75/200 column	Superdex 75/200 HR 10/30 column (GE Healthcare)
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
sec.	second
T_1	longitudinal /spin-lattice relaxation
T_2	transverse /spin-spin relaxation
T2SS	type two secretion system
T3SS	type three secretion system
T4P/T4PS	type IV pili/type IV pilus assembly system
TALOS	torsion angle likelihood obtained from shift and sequence similarity
TEM	transmission electron microscopy
TEMED	*N*,*N*,*N*′,*N*′-Tetramethylethylenediamine
UV	ultraviolet

Chapter 1 Introduction

1.1 Bacteria secretion systems

Bacterial secretion systems are responsible for the transport of proteins across the bacteria cell membrane. They are vital for the survival of the bacterium. Bacteria secrete a wide range of proteins from toxins and lytic enzymes to components of surface organelles such as flagella and pili. Bacteria secretion system also contribute to efflux pumps, which have a role in the colonization and the persistence of bacteria in the host and contribute to drug resistance (Piddock 2006).

1.1.1 Classification of bacteria

Bacteria can be classified according to their Gram staining. Gram-positive bacteria retain the dye crystal violet, which is removed with ethanol from the envelope of Gram-negative bacteria. The different staining properties are caused by different cell envelope structures. Gram-positive bacteria posses a single plasma membrane (also called the inner membrane (IM), Figure 1.1A) surrounded by a thick cell wall (CW) layer. Gram-negative bacteria have an additional membrane so that proteins secreted by Gram-negative bacteria must cross the periplasm and outer membrane (OM), in addition to the inner membrane (IM) (Figure 1.1B).

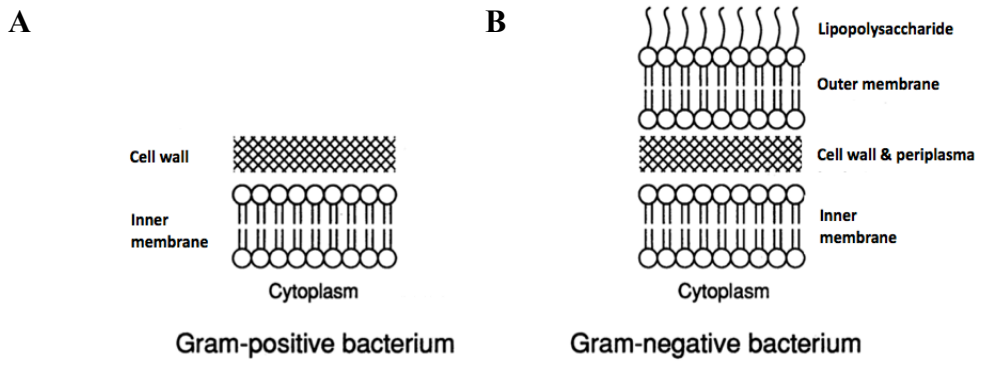


Figure1.1 Schematic representation of bacterium cell envelope. (A) Gram-positive and (B) Gram-negative bacterium(adapt from (Pugsley 1993).

In Gram-positive bacteria, secreted proteins are generally translocated across the single membrane by the Sec pathway or the twin-arginine (Tat) pathway. While Sec and Tat pathways are also used by the Gram-negative bacteria, more sophisticated secretion systems are present to facilitate transport across the outer cell envelope.

1.1.2 Architecture of the Gram-negative bacteria cell envelope

The space between inner and outer membrane of the Gram-negative bacteria is called the periplasmic space. Within the periplamic space, there is a layer of cell wall formed by peptidoglycan. This peptidoglycan layer is responsible for the cell's inability to retain the crystal violet stain upon decolorisation with ethanol during Gram staining. Both the inner and outer leaflets of the inner membrane are composed of phospholipids (Figure 1.2)(Tokuda 2009). Whereas, the outer membrane is an asymmetric bilayer possessing lipopolysaccharide (LPS) exclusively in the outer leaflet and phospholipid mostly in the inner leaflet (Mühlradt and Golecki 1975; Jaravine, Zhuravleva et al. 2008).

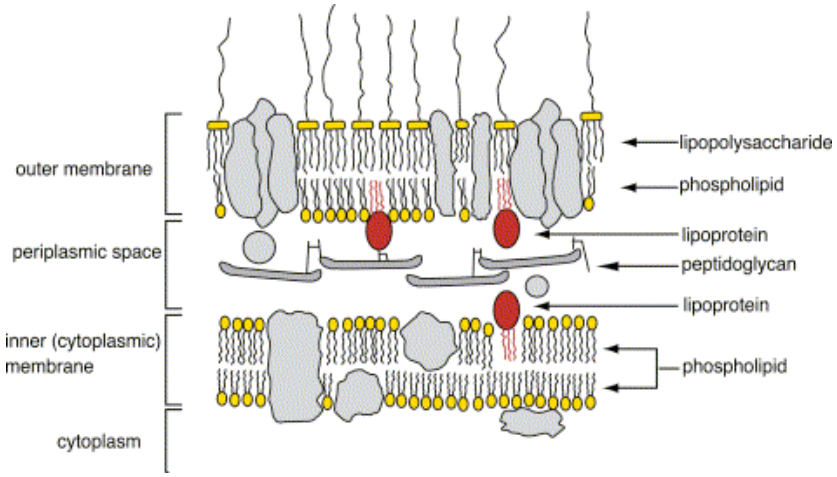


Figure 1.2 Gram-negative bacteria cell envelope structure. Note that the outer membrane is an asymmetrical bilayer containing LPS in the outer leaflet and phospholipids in the inner leaflet. At least 90 lipoproteins (indicated in red) are present in *E. coli*, most of which are specific to the outer membrane. Lipoproteins are anchored to the periplasmic surface of either the inner or outer membrane in *E. coli*. Figure adapted from (Tokuda and Matsuyama 2004)

The outer membrane functions as a permeability barrier to hydrophobic substances (Nikaido 2003). As a lipid biayer, the lipid portion of the IM and OM are impermeable hydrophilic molecules. But there is a class of channel forming proteins, called porins, located in the OM which provide a non-specific diffusion channel for transporting nutrients, such as amino acids, sugars and ions. The presence of porins means the periplasmic environment changes according to the surroundings.

However, the periplasm is a true cell compartment. It filled with periplasmic fluid, which has a gel-like consistency (Hobot, Carlemalm et al. 1984). Proteins located in the periplasm are all specifically targeted with a signal sequence.

1.1.3 Secretion in Gram-negative bacteria

Six different secretion pathways can facilitate protein secretion in Gram-negative bacteria (Thanassi and Hultgren 2000). Some secreted proteins are exported across the cell envelope in a single step via type I, type III, type IV or type VI pathways (Figure 1.3). Other secreted proteins are first transported to the periplasm via Sec or Tat pathways and then recognized and selectively translocated across the outer membrane via type II and type IV secretion system (Figure 1.3).

Most of the secreted proteins reach their final destination via the general secretory pathway (GSP). The type II secretion system (T2SS) is the major terminal branch of the GSP. Proteins secreted by T2SS include proteases, cellulases, pectinases, phospholipases, lipases and toxins. In general these proteins are associated with destruction of various host tissues, which help to supply nutrients to the bacteria leading to bacterial growth and disease.

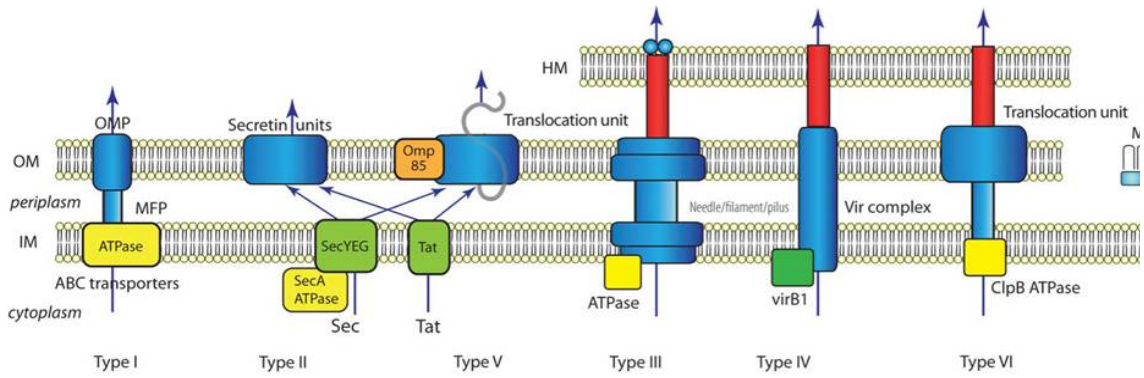


Figure 1.3 Secretion systems used by Gram-negative bacteria. In this simplified view only the basic elements of each secretion system are sketched. HM: Host membrane; OM: outer membrane; IM: inner membrane; MM: mycomembrane; OMP: outer membrane protein; MFP: membrane fusion protein. ATPases and chaperones are shown in yellow (Tseng, Tyler et al. 2009).

1.1.4 Type II secretion pathogens

The type II secretion system was first discovered in *Klebsiella oxytoca*, where it was shown to be required for secretion of the starch-hydrolyzing lipoprotein, pullulanase (d'Enfert C, Ryter A et al. 1987). Since then, this pathway has been found to be exploited by γ-proteobacteria, with a presence in representatives of 15 genera; *Acinetobacter*, *Aeromonas*, *Erwinia*, *Escherichia*, *Idiomarina*, *Klebsiella*, *Legionella*, *Methylococcus*, *Photobacterium*, *Pseudomonas*, *Shewanella*, *Vibrio*, *Xanthomonas*, *Xylella* and *Yersinia* (Sandkvist 2001; Peabody, Chung et al. 2003; Cianciotto 2005). It is now clear that T2S genes are also represented among the α-proteobacteria (e.g. *Bradyrhizobium*, *Caulobacter*, *Gluconacetobacter*, *Mesorhizobium*), the β-proteobacteria (e.g. *Azoarcus*, *Burkholderia*, *Chromobacterium*, *Ralstonia*) and the δ-proteobacteria (e.g. *Bdellovibrio bacteriovorus*, *Geobacter sulfurreducens*) (Cianciotto 2005) and ε-proteobacteria (*Bdellovibrio bacteriovirus, Geobacter sulfurreducens*). In the majority of listed organisms, genes for all of the essential T2SS components (Peabody, Chung et al. 2003) have been identified, which strongly suggests the presence of a functional T2SS in these bacteria.

1.1.5 Type II secretion related disease

There is considerable evidence for involvement of the T2SS in pathogenesis. First of all, the T2SS is conserved in many pathogens where it secretes degradative enzymes, which reveals that the system promotes the damage of host cells and tissues. In some instances, individual T2SS secreted proteins have been shown to contribute to virulence. Prominent examples include the cholera toxin (*V. cholerae*), heat labile toxin (*E.coli*) and exotoxin A (*P.aeruginosa*) (Tauschek, Gorrell et al. 2002).

Pathogens using T2SS as a major virulence pathway are responsible for a series of diseases (see reviews(Sandkvist 2001; Cianciotto 2005). In humans, these range from pneumonia (*L. pneumophila* and *P. aeruginosa*) to urinary tract infections (*E. coli*) and watery diarrhea (*V. cholerae*). In plants, soft rot disease is caused by *Erwinia* species.

In this thesis, the T2SS of *Erwinia chrysanthemi*, termed Out, is studied. *Erwinia chrysanthemi* is a soft rot pathogen, degrading succulent fleshy plant organs, such as roots, tubers, and stem cuttings, on a huge variety of economically important crops including potato, carrot and sugarcane. The tissue degrading diseases are caused by plant cell-wall degrading enzymes (e.g. pectate lyase, cellulase) secteted by *Erwinia* via the type II secretion system. The pathogen spreads easily under conditions of high humidity and free water. Disease development is highly dependent on high temperatures around 27-30^{o}C(Perombelon 1991). Nowadays, with the plastic packaging, global warming and the high food mileage, *Ewinia* is spreading all over the world and causes one of the most disastrous post-harvest losses to the farming industry(Perombelon 1991).

1.2 Type II secretion system components

There are about 12-15 genes identified as essential for T2SS function. The homologous genes and their subsequent proteins have been designated by letter A-O and S in most systems. Where in *Pseudomonas aeruginosa* and *Pseudomonas alcaligenes*, letters P-Z and A have been used instead.

In this thesis, Gsp is used to refer to general T2SS components. When referring to a component from a specific species, naming such as Out (*Ewinia*) and Pul (*Kelbsiella*) is used. When referring *Pseudomonas* component, the corresponding Gsp numbering component will be labeled as a subscript, such as *P.aeruginosa* GspD component is labeled as $XcpQ_D$.

T2SS proteins assemble to form a megadalton complex that spans the entire Gram-negative cell envelope. This complex includes a component in the cytoplasm (GspE), a secretin (GspD) and sometimes a pilotin (GspS) in the outer membrane and an inner membrane sub-complex that reaches into the periplasmic compartment and connects the cytoplasmic and outer membrane components (Sandkvist 2001).

Protein secretion is an energy consuming process. Cytoplasmic ATPase GspE provides energy for the whole system. GspE belongs to the large superfamily of "secretion superfamily ATPases". This superfamily involves the ATPase which functions in several marcomolecule transport systems, such as type II and type IV secretion. GspE is a Mg^{2+} dependent ATPase (Camberg and Sandkvist 2005; Yamagata and Tainer 2007) and exists as a hexamer in the functional T2SS (Yamagata and Tainer 2007). It is a hydrophilic protein, located in the cytoplasm when expressed alone in *E. coli*. It

achieves IM association through interacting with the IM protein GspL (Sandkvist, Bagdasarian et al. 1995).

On the inner membrane, GspC, GspF, GspL and GspM together with the periplasmic GspE are likely form a large homomultimer complex (Possot, Vignon et al. 2000). This inner membrane subcomplex is considered to interact with the outer membrane secretin GspD and pseudopilin GspG. GspM is likely to function as a dimer (Abendroth, Rice et al. 2004), known to localize and stabilize GspL in the complex (Michel, Bleves et al. 1998) and also appears to enhance the interaction between EpsE and EpsL (Sandkvist, Keith et al. 2000). OutF is found to co-immunoprecipitate with OutE and OutL. Co-purification of $XcpR_E$, $XcpS_F$ and $XcpY_L$ with histidine-tagged $XcpZ_M$ by metal affinity chromatography have also demonstrated interactions between the T2S components E_R, L_Y, M_Z and F_S. Possot et al. suggest that the Pul secreton components E, F, G, I, J, K, L, and M could all be replaced by the corresponding components of the Out secretons of *Erwinia chrysanthemi* and *Erwinia carotovora*, showing that they do not play a role in secretory protein recognition and secretion specificity (Possot, Vignon et al. 2000). GspC together with outermembrane secretin GspD, have been shown to involved in the specificity of the secretion system, and have been called the gatekeepers of the T2SS (Lindeberg, Salmond et al. 1996; Shevchik, Robert-Baudouy et al. 1997; Bouley, Condemine et al. 2001).

Pseudopilins GspG, H, I, J and K are believed to form a piston, which actively push substrate through the pore. The difference between pseudopilins is limited to the hydrophobic part of the N-terminal α-helix. GspG is the most abundant, so called 'major' pseudopilin forming majority of the pilus. Whereas GspH, I, J, K are the 'minor'

components considered involving in forming the 'tip" of pseudopilus (Korotkov and Hol 2008).

1.2.1 Structure of T2SS components

In past decade, there have been great advances in understanding molecular structure of T2SS components (Figure 1.4 and Figure 1.5). Many crystal structures of individual soluble domains have been solved in the lab of Professor Wim Hol (Abendroth, Rice et al. 2004; Abendroth, Murphy et al. 2005; Korotkov, Krumm et al. 2006; Korotkov and Hol 2008; Abendroth, Kreger et al. 2009; Abendroth, Mitchell et al. 2009; Korotkov, Pardon et al. 2009). The outer membrane secretin has also been studied by transmission electron microscopy (TEM) (Nouwen, Ranson et al. 1999; Reichow, Korotkov et al. 2010).

At the start of this work two key unknown structures were GspC homology region (GspC-HR) and the secretin's pilotin, GspS. The structures of the OutC-HR and the GspS are discussed in this Thesis.

1.2.1 Regulation of T2SS

The T2SS components as well as the expression of genes of secreted proteins are regulated by quorum sensing (QS) (Wooldridge 2009; Goo, Kang et al. 2010), which uses acylated homoserine lactones as autoinducers in Gram-negative bacteria (Miller and Bassler 2001). The small molecule autoinducer accumulates and when the cellular threshold is reached, they activate translational regulators. In *P. aeruginosa*, it is found that the expression of *gspE-M* is induced earlier than the *gspCD* operon. In *E. carotovora*, T2SS secreted protein Svx is also found to be regulated by QS in

transcription level (Corbett, Virtue et al. 2005)

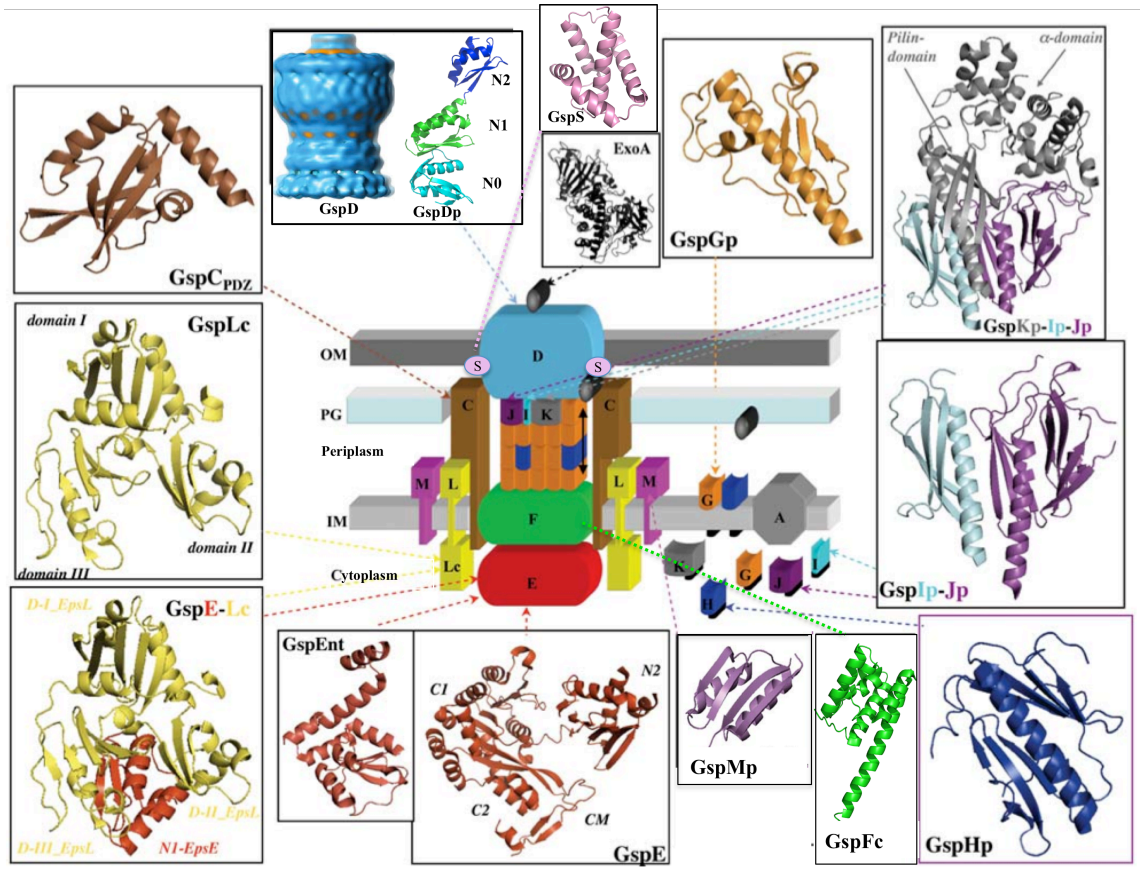


Figure 1.4 Model of the T2SS and known high-resolution conponent structures. Names refer to the Gsp nomenclature. In the central schematic diagram, the single-letter code is used and refers to the Gsp nomenclature (i.e., GspD=D). Inserts show the high-resolution structures of individual components and sub-complexes resolved so far. The crystal structures are shown in ribbons representation using Pymol (http://pymol.sourceforge.net). The colour code matches model with structure for each components. The PDB codes are: GspCPDZ (2I6V); GspD-Nter (3EZJ), GspE (1P9R); GspEnt (2D27); GspE-L (2BH1); GspFc(2VMA); GspG (1T92); GspH (2QV8); GspI-J (2RET); GspK-I-J (3CI0); GspL (1W97); GspM (1UV7); ExoA (1IKQ). For the secretin GspD, the EM map (EMDB-1763) has been used. Outer membrane (OM); Inner-membrane (IM); Peptidoglycan (PG). Suffixe "p"and "c" indicates a "periplasmic" and "cytoplasmic" domain respectively. (Durand, Verger et al. 2009)

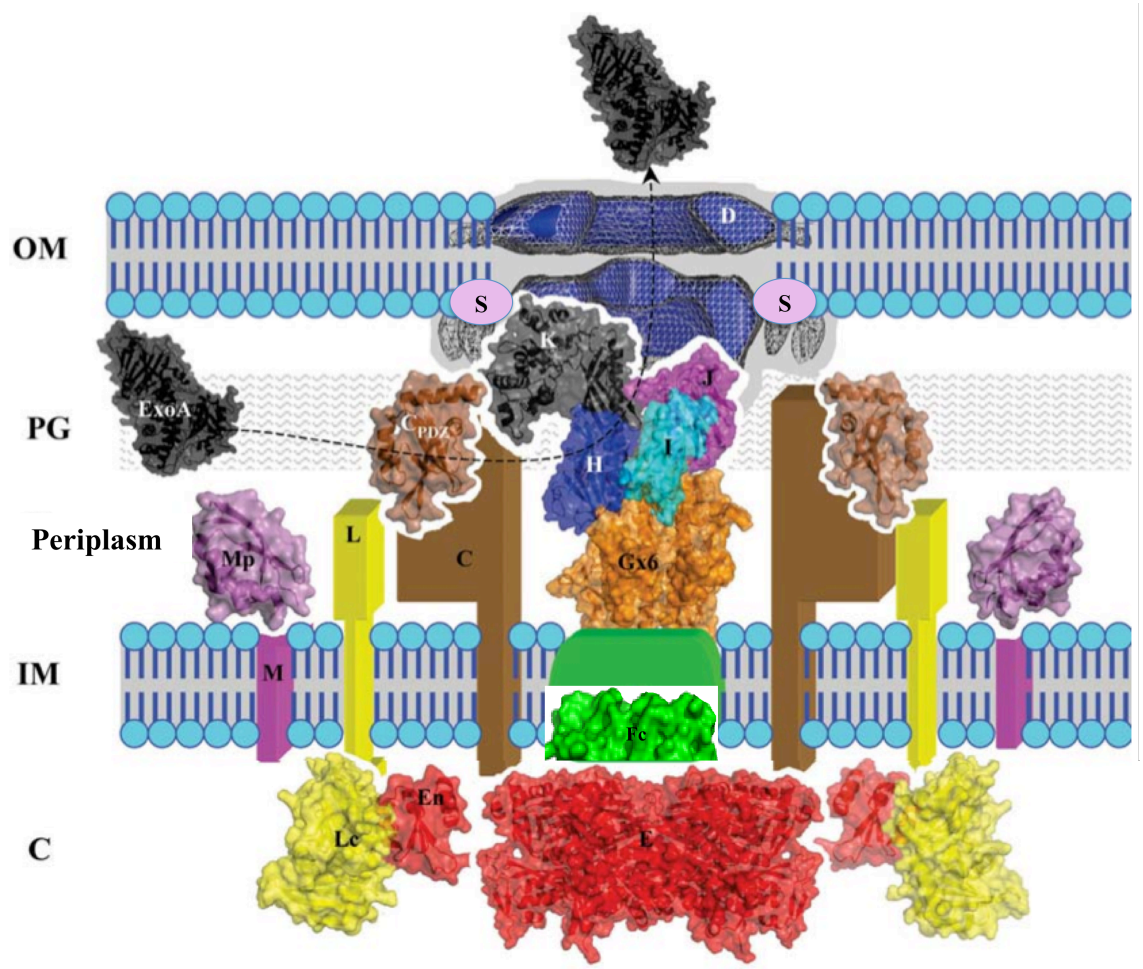


Figure 1.5 High-resolution model of the T2SS. All the known high-resolution structures of individual components and sub- complexes are represented according to the same color-code used in Figure 1.2.1. The crystal structures are shown in ribbons representation and transparent surface using PyMOL (http://pymol.sourceforge.net). Only the domains of known structures have been represented, with a cartoon for the rest of the protein. The model has been built to fit with the most recent biological and structural data on T2SS secretons. For GspE, the hexameric model of EpsE is shown. The pseudopilus structure is arbitrary composed of 6xGspG, 1xGspH, 1xGspI, 1xGspJ, and 1xGspK subunits that are packed together in a similar arrangement as the pilin subunits in type IV pili (Durand, Verger et al. 2009)

1.3 Secretins

Secretins are a class of outer membrane β-barrel proteins, with subunit mass ranging from 50-70 kDa in size. The secretin subunits associate into a large and highly stable multimeric protein conducting channel in the OM of Gram-negative bacteria (Lindeberg, Salmond et al. 1996; Burghout, Beckers et al. 2004). Type II secretion system (T2SS), type III secretion system (T3SS) and the type IV pili system (T4PS) are the protein secretion systems features secretin (Figure 1.3.1). Secretin is also found in participating in the assembly and extrusion of filamentous bacteriophages. In *Vibro cholera*e, the secretin EspD, function in both T2SS and bacteriophage (Davis, Lawson et al. 2000).

1.3.1 Structure of secretins

Secretin subunits form an outer membrane pore of 50-80 Å in width which provides a channel for the translocation of folded proteins, assembly of oligomeric pilins or for DNA uptake. The opening and closing of the pore is carefully regulated to maintain the integrity of the outer membrane and the periplasmic contents.

The domain composition of secretins is shown in figure 1.3.2A. The C-terminal half of the protein contains the secretin domain, which consists largely of TM β-strands forming a 10 to 13 β-strands β-barrel structure in the OM. Whereas the N-terminus exhibits conservation within subgroups from related transport pathways (Genin and Boucher 1994) . The secretin N-terminus is thought to be largely free of transmembrane regions and to reside in the periplasm (Nouwen, Stahlberg et al. 2000). The extreme C terminus of the GspD, which includes the pilotin interaction sequence is also considered to reside in the periplasm (Daefler, Guilvout et al. 1997).

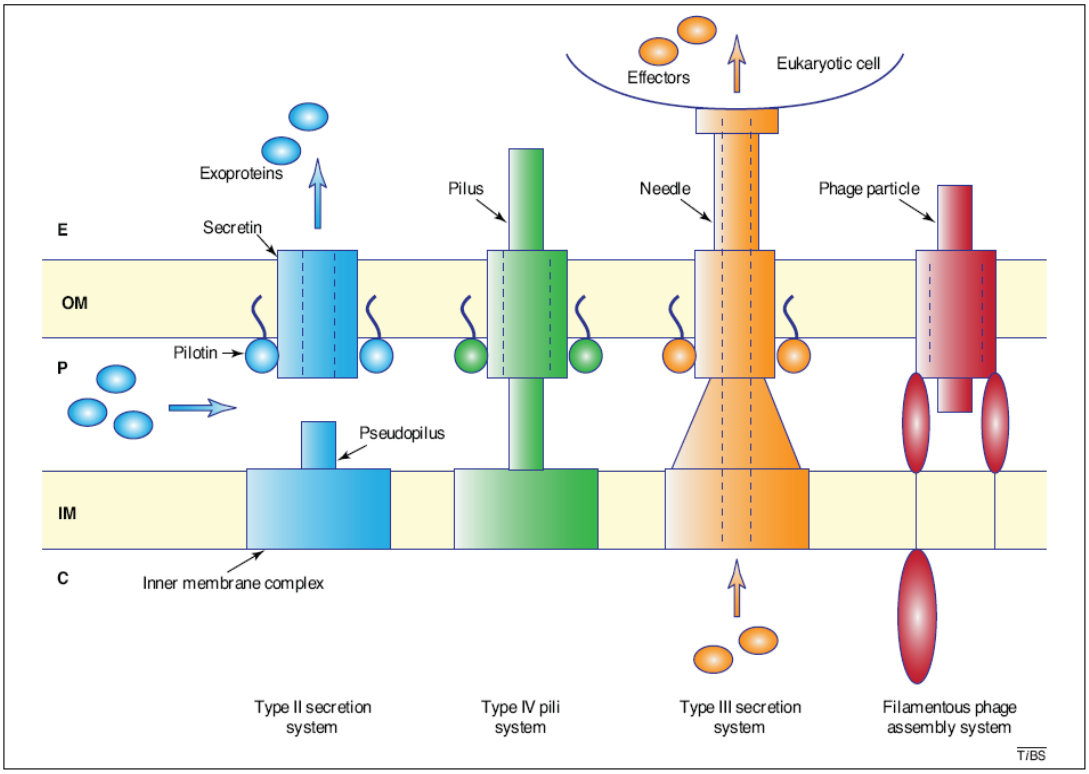


Figure 1.6 Secretins in Gram-negative bacteria. Schematic view of the type II and type III secretion systems, type IV pili system, and bacteriophage assembly system. The secretin is the major outer-membrane component of all these systems. The insertion of secretins into the outer membrane is often assisted by specific lipoproteins called pilotins. The T2SS secretes exoproteins from the periplasm to the extracellular space in the folded form. The T2SS pseudopilus is formed by multiple pseudopilin subunits, and the pseudopilus is thought to act as a piston and/or plug during the secretion process. The T4PS is related to the T2SS in several architectural and functional aspects, but a key difference is that the pilus extends outside the bacterial surface. The T3SS transport effectors directly to the eukaryotic cytoplasm or membrane via a hollow needle. The inner-membrane complexes of the T2SS, T4PS and the T3SS are composed of multiple proteins that include at least one ATPase, which is involved in providing energy for secretion or pilus extension/retraction processes. The filamentous phage-assembly system is composed of a secretin and two inner-membrane proteins. C, cytoplasm; E, extracellular space; IM, inner membrane; OM, outer membrane; P, periplasm. (Korotkov, Gonen et al. 2011)

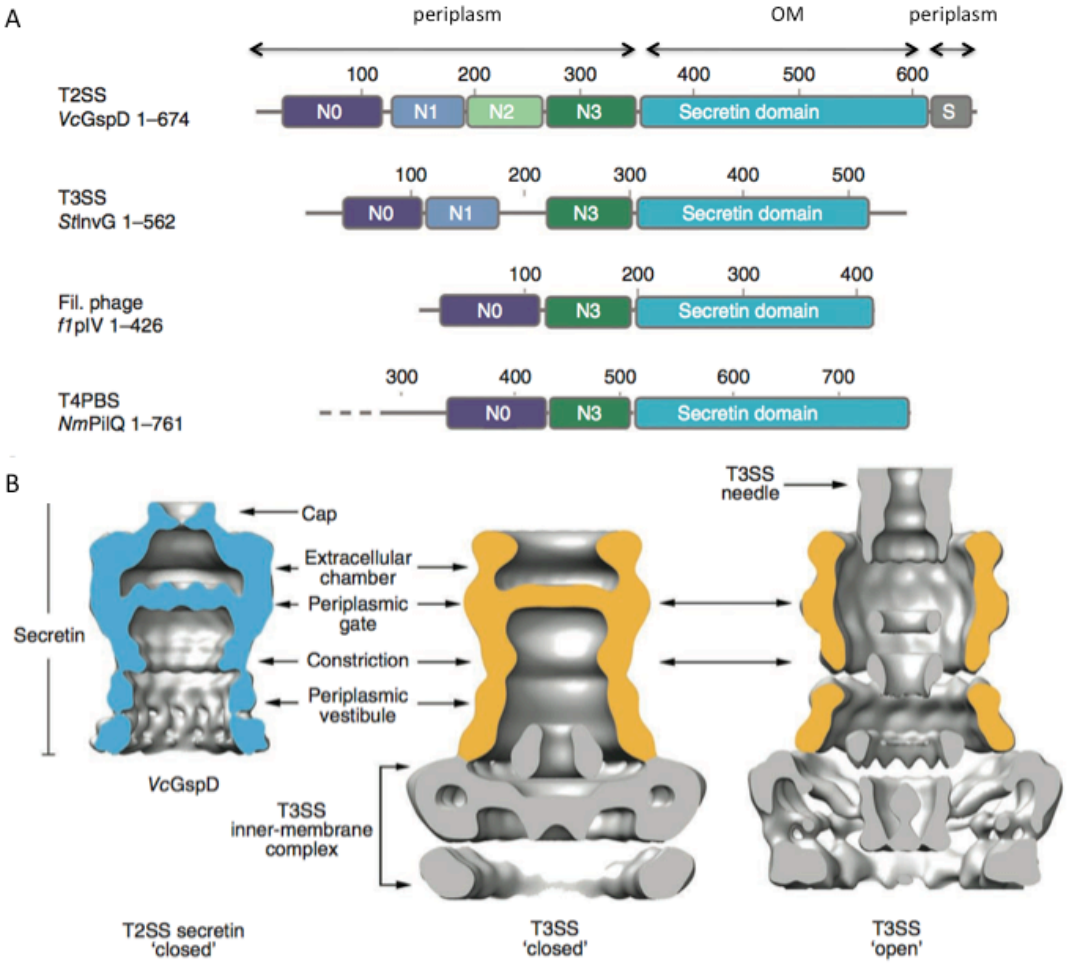


Figure 1.7 The secretin architecture is conserved in different secretion systems. (A) Domain architecture of secretins from the T2SS, the T3SS, the filamentous phage assembly system and the T4PBS. Members of the secretin superfamily contain a C-terminal secretin core homology domain17,18 (cyan). The T2SS secretins generally contain four periplasmic subdomains, termed N0–N3. The N0 subdomain (dark blue) is located at the N terminus and is followed by the three structurally homologous subdomains, N1–N3 (blue, green and dark green, respectively).A T2SS-specific domain, termed the S-domain (gray), is located at the very C terminus. Secretins from other systems share a similar architecture, composed of the secretin domain and at least two periplasmic subdomains that are structurally equivalent to N0 and N3 of VcGspD. (B) Structural comparison of the VcGspD density (blue, left) to single-particle reconstructions of the T3SS in its closed state (middle) (EMDB 1224) and to the fully assembled T3SS needle complex in its open state (right) (EMDB 1617). The outer- membrane T3SS secretin sits on top of a large inner-membrane complex (gold, middle and right). VcGspD seems to be in its closed state (compare left panel with middle panel). (Reichow, Korotkov et al. 2010).

Several secretins have been imaged by TEM. A ring-shaped structure with large central channel was revealed (Figure1.7) (Koster, Bitter et al. 1997; Linderoth, Simon et al. 1997). The secretin pore is a large structure in the OM, ranging from 12-20nm in diameter, with central channel ranging from 5-10 nm in diameter. The secretins in different secretin systems all appear to be roughly cylindrical in shape.

The channel forming and oligomerisation activities appear to locate on the C-terminal secretin domain. The secretin domain of the secretin PulD forms heat resistant dodecameric complexes within less than 10 minutes in the *in vitro* transcription-translation system containing liposome (Guilvout, Nickerson et al.).

1.3.2 T2SS secretin

GspD, the secretin, in T2SS is one of the non-interchangeable subunits in the T2SS and is the only integral outer-membrane component (Hardie, Lory et al. 1996). It exists as a 12-14 subunit oligomer. Each subunit consists of a periplasmic N-terminal domain, which includes N0, N1, N2 and N3 subdomains, a transmembrane secretin domain, and a periplasmic pilotin interacting S domain.

The C-terminal secretin domain is the most conserved domain forming the transmembrane pore and is responsible for oligomerisation (Guilvout, Nickerson et al.). At the time of writing, there is no secretin domain structure available, but it is predicted to comprise TM β-strands, forming a β-barrel structure (Pugsley 1993).

The secretin N-terminal domain is distinct between different secretion systems. It extends into periplasm and is involved in a series of interactions with secreted protein as

well as other T2SS components. GspD was shown to interact with secreted protein (Lindeberg, Salmond et al. 1996; Shevchik, Robert-Baudouy et al. 1997; Bouley, Condemine et al. 2001) as well as the HR of GspC (Korotkov, Krumm et al. 2006; Login, Fries et al. 2010). Yeast two-hybrid studies shown GspJ interacts with the periplasmic part of GspD (Douet, Loiseau et al. 2004). GspB has only been identified in a few species, in these GspB and GspD stabilized each other. GspB co-fractionates with then OM and can be cross-linked into a larger complex in the presence of GspD (Condemine and Shevchik 2000).

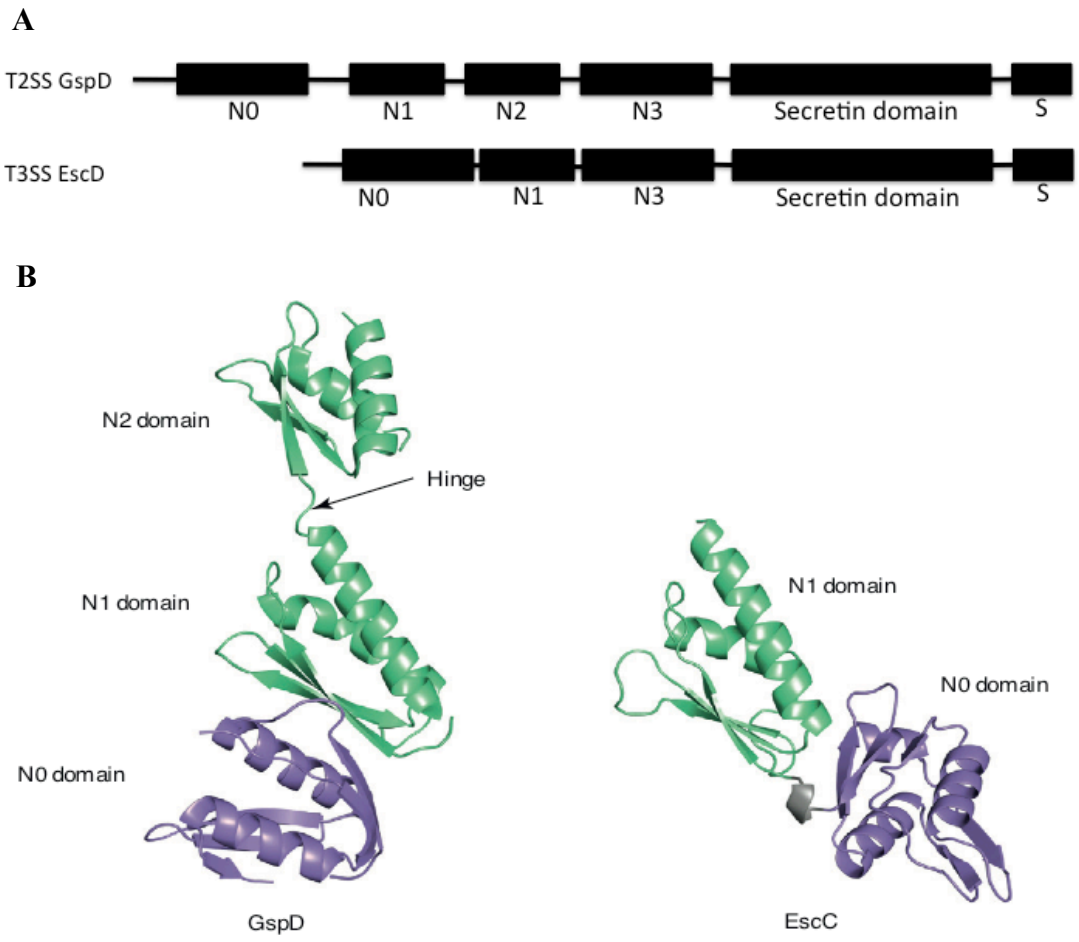


Figure 1.8 GspD modular organization and structure. (A) Domain composition of secretin. (B) Crystal structure of secretin N-terminal domains. The N0, N1 and N2 structure of T2SS GspD and N0, N1 stucture of T3SS EscC. (Korotkov, Gonen et al. 2011).

The crystal structure of ETEC GspC N-terminus, including N0, N1 and N2 domains, was solved by X-ray crystallography in the presence of a nanobody (Korotkov, Pardon et al. 2009). The N0 domain reveals a structure similar to the TonB-dependent receptor (Figure 1.8). N1, N2 domains reveal a fold similar to the eukaryotic type I KH domain, which are typically involved in binding RNA and DNA. However, the KH domain positively charged RNA/DNA binding residues are absence in N1 and N2. The N3 domain structure is not known, but over 15% sequence identity between N1 N3 and N2 N3 suggests N3 is likely to have a KH domain fold as N1 and N2. It is interesting to note that he N0 and N1 domain of type III secretin crystal structure revealed a different orientation as the type II one (Figure 1.8B). The current GspD N-terminal domain structure is solved in the presence of nanobody, which binds both N0 and N1 and could have a substantial influence on the domain orientation.

The S-domain of GspD is responsible for the direct interaction with GspS (Shevchik and Condemine 1998), which stabilizes GspD and is responsible for the transport of GspD to the OM (Daefler, Guilvout et al. 1997; Shevchik, Robert-Baudouy et al. 1997). The transportation from the periplasm to outer membrane is achieved by the Lol pathway with the help of a chaperone pilotin (Collin, Guilvout et al. 2011). The secretin is transported from periplasm to OM as individual subunits (Collin, Guilvout et al. 2011).

Several EM structures of GspD are available that give a 3D overview of oligomeric secretin. The detailed structure of the C domain and the S domain have yet to be determined, and the arrangement between the individual sub-domains and domains is not yet clear.

1.4 Type II secretin and its partners

The T2SS secretin, GspD forms large homomultimeric structures in the outer membrane. The secretin adopts a ring shaped structure and forms a pore, through which secreted protein pass (Nouwen, Ranson et al. 1999; Chami, Guilvout et al. 2005; Reichow, Korotkov et al. 2010). Precise regulation of this channel is crucial, since an open pore can lead to leaking of the periplasmic contents. The localization and regulation of the secretin is mainly achieved through interaction with two of its partners. The regulation of the secretin is mainly controlled by its interaction with the inner membrane GspC (Lee, Wang et al. 2000; Lee, Chen et al. 2004). The localization of the secretin is facilitated by its pilotin GspS.

1.4.1 GspC and GspD; gatekeepers of the T2SS

GspC and GspD are the gatekeepers of the T2SS. In search of the gatekeeper of T2SS, Lindeberg et al. created a non-polar mutation of each of *E. chrysanthemi* gene in the OutC-M and OutS and OutB operon and attempted to complement the gene with its *E. carotovora* homologue (Lindeberg, Salmond et al. 1996). They found that all of the *E. carotovora* T2SS proteins could substitute for *E. chrysanthemi* proteins except for OutD and OutC.

Later, Possot and colleagues used a similar technique and found all Pul genes except pulC and pulD genes could be exchanged with the *Erwinia* homologues without disrupting pullulanase secretion (Possot et al., 2000). Bouley et al suggest that the PDZ domain of GspC and N-terminal GspD are required for different secreted protein recognition (Bouley, Condemine et al. 2001). It has also been suggested that GspC and GspD could be interacting to dock the inner membrane subcomplex with outer

membrane secretin pore (Lindeberg, Salmond et al. 1996; Shevchik, Robert-Baudouy et al. 1997).

1.4.2 GspC is a key player in T2SS

GspC locates on the inner membrane and plays a key functional role in T2SS. It interacts with secreted proteins (Bouley, Condemine et al. 2001) as well as the outer-membrane secretin pore (Lindeberg, Salmond et al. 1996; Shevchik, Robert-Baudouy et al. 1997). The interaction between GspC and GspD provides the structural and functional integrity of the secretion machinery across the two cell membranes. The interaction with the secreted protein guides the secreted protein to the section system and through the pore in the outer membrane. With the hydrolysis of ATP, secreted protein is pushed through the secretin pore to the external medium (Figure 1.9A).

GspC is a transmembrane protein, which spans the inner membrane. It consist of a short N-terminal cytoplasmic sequence, a transmembrane segment (TMS) and a large periplasmic region (Figure 1.9B). On the periplasmic region, there are two defined regions. One is the homology region (HR), which is conserved across species and the other domain is involved in protein-protein interaction, which is either a PDZ domain (in *Vibrio, Erwinia, Escherichia* and others) or a coiled-coil domain (in *Pseudomonas)*.

GspC has been shown to function as a dimer, where the TMS is responsible for dimer formation (Login and Shevchik 2006). Periplasmic targeting and dimerisation are crucial for the assembly of a functioning secretion system *in vivo* (Login and Shevchik 2006).

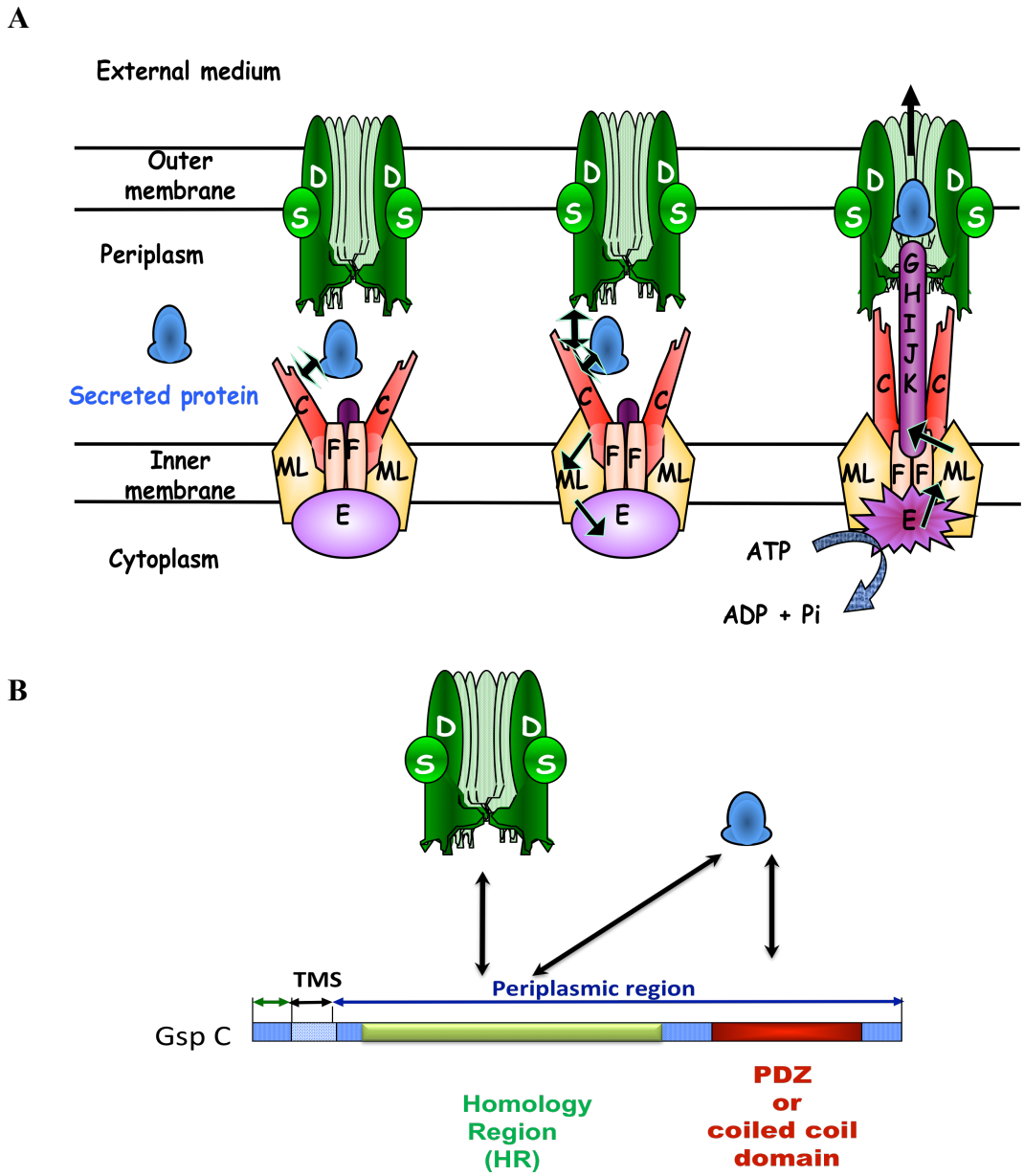


Figure 1.9 Schematic diagram showing the role of GspC. (A) Current understanding of GspC involves in interactions with GspD and secreted protein within the T2SS. (B) GspC subdomains GspC interact with Exo-protein as well as secretin formed by GspD in the periplasma. The docking between GspC, GspD and exo-protein activate the T2SS, stimulating pilus formation and protein secretion.

The PDZ domain interacts with secreted proteins. Functional analysis of truncated OutC variants, carrying deletions in the periplasmic domain, revealed the PDZ domain is responsible for secretion specificity of some but not all proteins (Bouley et al. 2001). Login et al. showed that the GspC PDZ region does not interact with GspD (Login,

Fries et al. 2010). GspC PDZ structure of *Vibrio cholerae* (Korotkov, Krumm et al. 2006) and *Erwinia chrysanthemi* (work of Drs.Fries and Hutchison in my Lab.) have been solved by crystallography.

The homology region of GspC, GspC-HR, is essential for T2SS function. *C-terminal truncated* $XcpP_C$ (GspC of *Pseudomonas aeruginosa), without the coiled-coil domain was able to complement XcpP deletion (Bleves, Gerard-Vincent et al. 1999). Similarly, GspC of Xanthomonas campestris* (XpsN) was able to complement an *xpsN* deletion mutant (Lee, Chen et al. 2004). This particular C-terminal truncation (XpsNΔ159) ends exactly at the end of predicted HR domain. A two extra residue truncation (XpsNΔ157) was however inactive. In *Erwinia chrysanthemi* PDZ-less GspC variants, full activity was observed for a subset of secreted protein (Bouley, Condemine et al. 2001).

Compared to the PDZ domain, there is limited structural knowledge about GspC-HR. J-pred secondary structure prediction of GspC-HR suggested a beta-sheet dominated structure (http://www.compbio.dundee.ac.uk/www-jpred/). Hence, the initial motivation of this study was to solve the structure of GspC-HR.

1.5 Secretin's pilotin

The interaction between secretin GspD and its lipoprotein pilotin is essential for secretin translocation in several species. Common to T2SS, T3SS and Type IV pili is the observation that proper targeting of secretin to the outer membrane requires the presence of pilotin (Shevchik and Condemine 1998; Schuch and Maurelli 2001; Guilvout, Chami et al. 2006). The known examples pilotins are PulS and OutS of the T2SS of *Klebsiella* and *Erwinia*, InvH, MxiM and YscW of the T3SS of *Salmonella*,

Shigella and *Yersinia* and PilW and PilF of the T4P of *Neisseria* and *Pseudomonas* (Hardie, Seydel et al. 1996; Crago and Koronakis 1998; Shevchik and Condemine 1998; Burghout, Beckers et al. 2004; Lario, Pfuetzner et al. 2005; Koo J 2008).

1.5.1 Pilotin lipidation

Pilotin has a lipid-modified cysteine residue close to its amino termini through which the protein is anchored to the membrane. Lipoprotein precursors are sequentially processed to their mature forms on the periplasmic side of the inner membrane after their translocation across the inner membrane by Sec translocon (Hayashi and Wu 1990; Pugsley 1993). Analysis of signal sequences of 26 distinct lipoprotein precursors has revealed a consensus sequence of lipoprotein modification/processing site of Leu-(Ala, Ser)-(Gly, Ala)-Cys at – 3 to + 1 positions which would represent the cleavage region of about 75% of all lipoprotein signal sequences in bacteria. Unmodified pro-lipoprotein with the putative consensus sequence undergoes sequential modification and processing reactions catalyzed by glyceryl transferase, O-acyl transferase(s), pro-lipoprotein signal peptidase (signal peptidase II), and N-acyl transferase to form mature lipoprotein (reviewed in (Tokuda and Matsuyama 2004).

In *E. coli* lipoproteins are anchored to the periplasmic side of either inner or outer membrane through N-terminal lipids depending on the lipoprotein-sorting signal (Pugsley 1993). Yamaguchi et al. first reviewed the importance of the residue at position 2 for the sorting of E.coli lipoprotein (Yamaguchi, Yu et al. 1988). They showed that replacement of Ser at position 2 of an outer-membrane-specific lipoprotein by Asp caused the protein to remain in the inner membrane. Furthermore, replacement of Asp at position 2 of an inner membrane-specific lipoprotein by another residue

caused outer membrane localization. For the T2SS pilotin, GspS, position 2 is not Asp (Figure 1.10) showing it locates on the periplasmic side of outer membrane.

The attached acyl lipid is essential for Lol pathway shuttle chaperone protein (LolA) recognition. In T3SS, an unlipidated form of the YscW protein was not functional, although it still interacted with the secretin and caused mislocalization of YscC even in the presence of wild-type YscW (Burghout, Beckers et al. 2004). It has also been shown that the sorting function of PulS can be replaced by fatty acylation of the PulD N-terminus (Collin, Guilvout et al. 2011).

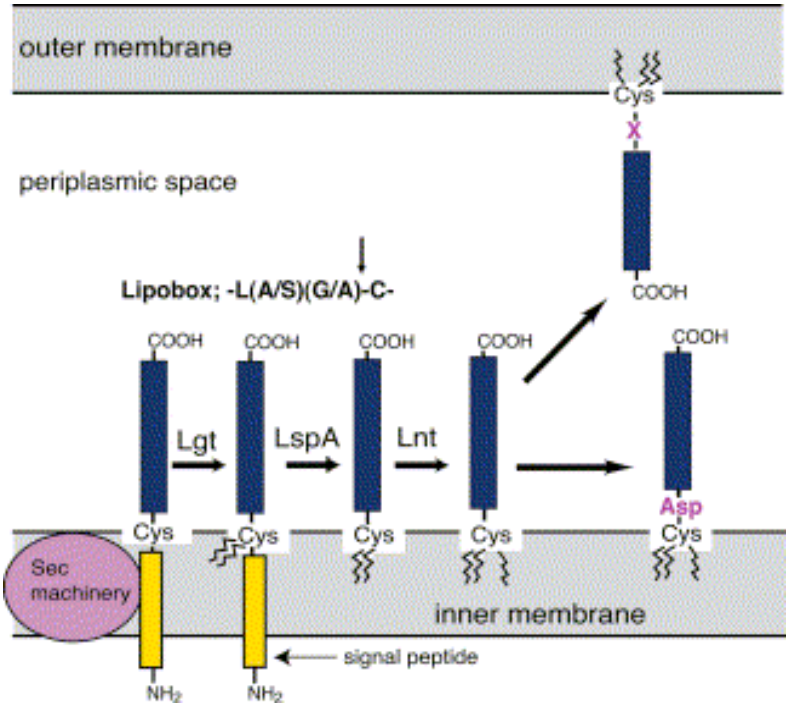


Figure 1.10 Biogenesis of lipoproteins. Consensus lipobox sequences are indicated with a signal cleavage site indicated by an arrow. X represents a residue other than Asp. Lgt, phosphatidylglycerol/prolipoprotein diacylglyceryl transferase; LspA, prolipoprotein signal peptidase (also called Spase II); Lnt, phospholipid/apolipoprotein transacylase. (Tokuda and Matsuyama 2004).

1.5.2 Lol system

The Lol system, composed of five Lol proteins, is responsible for the transport of lipoproteins to the outer membrane (Tokuda and Matsuyama 2004; Collin, Guilvout et al. 2011). All five Lol proteins, A to E, are highly conserved in various Gram-negative bacteria. The system comprises LolCDE complex in the innermembrane, LolA in the periplasm and LolB in the outer membrane (Figure 1.11).

LolA is a shuttle chaperone. It recognizes the acyl chains on lipoproteins (Tokuda and Matsuyama 2004; Lario, Pfuetzner et al. 2005). Lipoproteins released from periplasm and in the presence of LolA were found to exist as water-soluble complexes with LolA. Therefore, forming the complex with LolA allows lipoproteins to travel between the inner and outer membranes. Recently, the PulS-LolA complex was confirming that GspS is transported by Lol system (Collin, Guilvout et al. 2011).

LolCDE, is an ABC transporter which releases lipoproteins from the inner membrane driven by ATP hydrolysis (Yakushi, Yokota et al. 1998).

LolB is a lipoprotein on the outer membrane. Although there is no apparent homology between amino acid sequence of LolA and LolB, their structures are strikingly similar (Takeda, Miyatake et al. 2003). Lipoproteins are transferred from LolA to LolB due to the affinity difference between their hydrophobic cavities for lipoprotein lipid. LolB transfers the associated lipoproteins to the inner leaflet of outer membrane to which lipoproteins are anchored through three acyl chains (Tokuda and Matsuyama 2004).

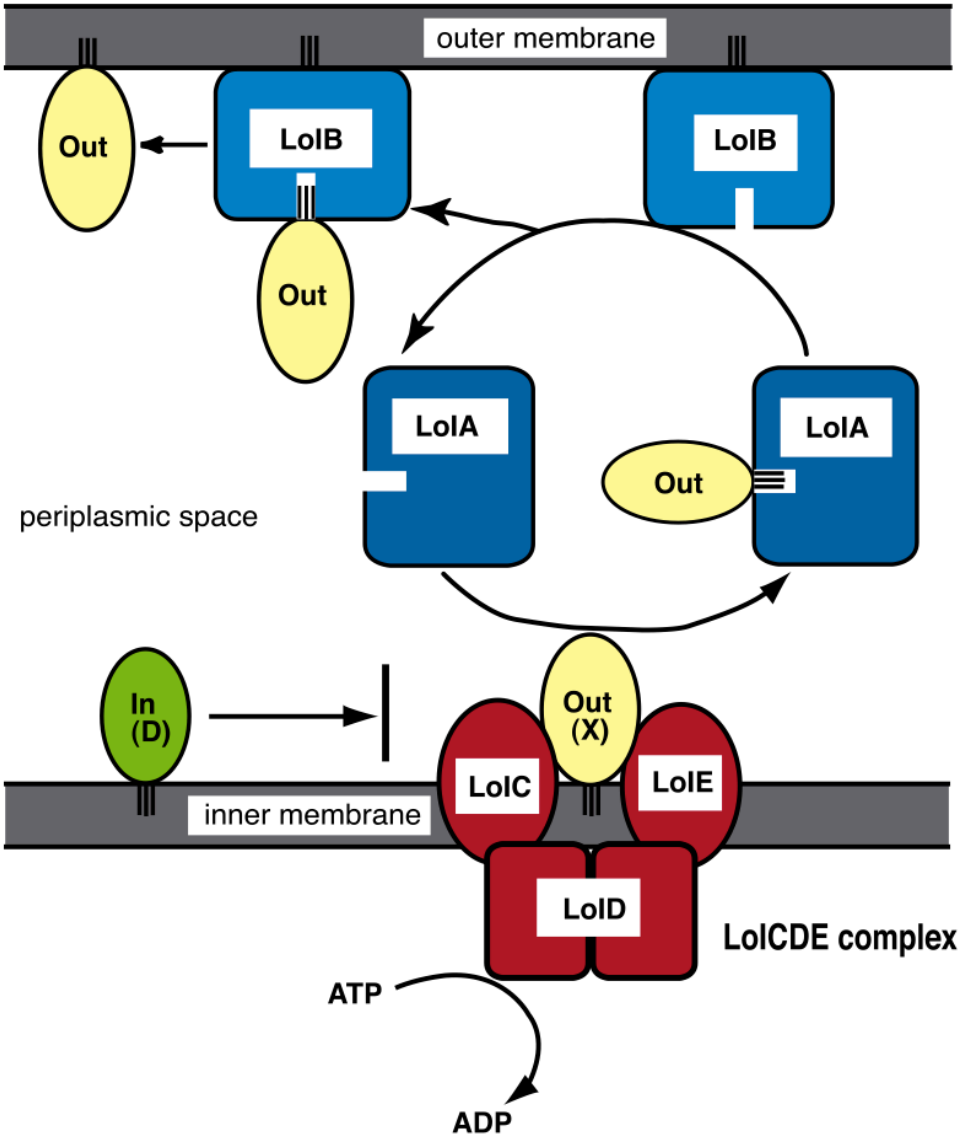


Figure 1.11 Sorting and outer membrane localization of lipoproteins by the Lol system. ‘‘In’’ and ‘‘Out’’ represent inner membrane-specific and outer membrane-specific lipoproteins, respectively. An ABC transporter, LolCDE, releases outer membrane-specific lipoproteins from the inner membrane, causing the formation of a complex between the released lipoproteins and the periplasmic molecular chaperone LolA. When this complex interacts with outer membrane receptor LolB, the lipoproteins are transferred from LolA to LolB and then localized to the outer membrane. The inner membrane retention signal Asp at position 2 inhibits the recognition of lipoproteins by LolCDE, thereby causing their retention in the inner membrane. (Tokuda and Matsuyama 2004)

1.5.3 Structure of pilotin

Before this work the known pilotin structures were MxiM of the T3SS of *Shigella flexneri* and PilW/PilF(Lario, Pfuetzner et al. 2005) of the T4PS of *Neisseria meningitides*/ *Pseudomonas aeruginosa*.

It is interesting to note that the T2SS, T3SS and T4PS secretins's pilotin all share similar function but have striking different folds (Figure. 1.12).

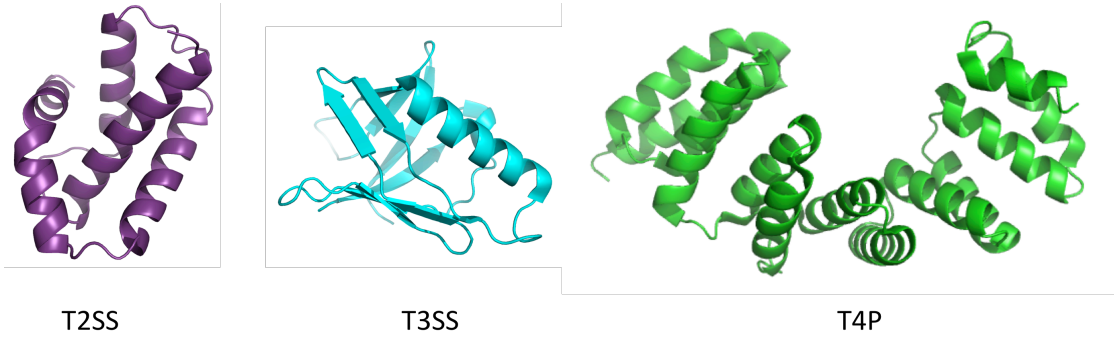


Figure 1.12 Pilotins in Gram-negative bacteria. Ribbon representation of the T2SS, T3SS and T4P pilotin. The pilotin crystal structures are shown in ribbon representation. T2SS pilotin GspS (our Lab.), T3SS MxiM (PDB:1Y9L) and T4P pilotin PilF/W(PDB: 2FI7) were selected as represent the type of fold.

The structure of the T2SS pilotin, GspS, revealed a nested α-helix hairpin domain. Whereas, T4PS pilotin, PilF/W, consists 13 helices, which form a right handed TPR superhelix (Kim, Oh et al. 2006). The cracked β-barrel structure of T3SS pilotin, MxiM, has been solved in complex with an 18 residue peptide from the cognate secretin MxiD and the authors propose a model for the way MxiM assists MxiD assembly (Okon, Moraes et al. 2008).

1.6 Aims of this study

The wide spread emergence of antibiotic resistance in pathogenic bacteria requires the exploration of novel classes of antibacterial compounds. For virulence spreading

through T2SS, T3SS and T4PS, secretins offer the only passage for secreted protein or virulence factor to cross OM. Hence the secretin could be an attractive target for intervention using drugs.

From this Chapter, it is clear that although recently there have been great advances in understanding the molecular structure of the T2SS, the molecular structure of GspC-HR, an essential component of T2SS gatekeeper GspC and responsible for interaction between GspC and GspD, is unknown. There is also relatively little knowledge about the key interactions involved in assembly, stability and regulation of the T2SS at the molecular level.

Determination of the structure of GspC-HR was the initial aim of the study. The interaction between GspD and GspC-HR is a key event in the formation of a functional T2SS. In the first results chapter of this Thesis I describe how the GspC-HR structure was solved by NMR spectroscopy (Chapter 4). Production of soluble GspD N terminal domains and the interaction between GspC-HR and GspD-N0 is then described (Chapter 5).

Since the GspS structure was solved in the lab recently in the lab, revealed a novel pilotin fold. The pilotin secretin interaction became most interesting. It was suggested that pilotin, OutS interacting with OutD C-terminal 62 amino acids(Shevchik and Condemine 1998). In Chapter 7, the interaction between OutS and OutD is further characterized.

Chapter 2 Material and methods

2.1 Molecular biology

2.1.1 Plasmid construction

pGEX-6P-1 vector (GE Healthcare) containing a OutC-Peri, OutC-HR and OutD-Cter; pET20b (GE Healthcare) containing OutD-N0, OutD-N0N1, OutD-N1N2, OutD-N1, OutD-Peri and OutS were provided by our collaborator Dr. Vladimir Shevchik (Lyon). OutC-HR and OutD-Peri were also used as the template from which truncated constructs were generated.

2.1.2 Bacterial strains

E. coli alpha select competent cells (Bioline) were used for initial cloning and DNA purifications procedures. *E. coli* BL21 (DE3) competent cells (Bioline) used in expression of all protein constructs. The DE3 lysogen contains the T7 polymerase gene under the control of the *LacUV5* promoter. Addition of IPTG (isopropyl-b-D-thiogalactopyranoside) (Fisher) induces constitutive expression of T7 polymerase. For Expression vectors under the control of a T7 promoter, the polymerase transcribes the mRNA at high copy number resulting in over-expression of the desired protein.

2.1.3 DNA manipulation and analysis

Agarose gel were prepared and run according to the standard procedures (Sambrook 2001) to separate DNA according to the size. All plasmid DNA was purified from overnight culture of E.coli alpha-select competent cells by peq GOLD Mini prep (peq lab), according to manufacturer's instruction. Final DNA concentration was detected by UV spectrophotometer according to the following:

$$\text{Concentration } (\mu g/ml) = (A_{260} - A_{320}) \times 50 \mu g/ml$$

Where A_{260} reading is where DNA absorbs light most strongly and A_{320} is measurement for turbidity. DNA purity is estimated by A_{260}/A_{280} ratio. An A_{260}/A_{280} ratio between 1.7 and 2.0 generally represents a high-quality DNA sample.

2.1.4 Polymerase chain reaction (PCR)

PCR were carried out using Microcycler (MJ Research) and the Pfu polymerase (Novagen). All reactions were carried out in thin wall 200µl PCR tubes (Starlab) at the final volumn of 50ul. A typical PCR reaction setup is shown in Table 2.1 and a typical PCR reaction program setup is shown in Table 2.2.

Reagent	Volumn (µl)
Pfu Reaction Buffer (10x)	5
dNTP (2.5mM)	4
Template DNA	1
Forward Primer	1
Reverse Primer	1
dH_2O	37
Pfu polymerase	1
Total	50

Table 2.1 Typical PCR reaction mixture. Where possible primers were designed to conform optimized parameters: GC content 40-60%, length 25-35 bases, G or C at the end of the primer.

Step	Temperature (^oC)	Time (Sec)
1	95	120
2	95	30
3	55	60
4	72	60
5	Cycle step 2-4 30 times	
6	72	1min
7	4	5min

Table 2.2 Typical thermocycler program for PCR reaction. An initial 2min melt to ensure full priming in the first cycle. In most cases 55^{o}C was found to be sufficient regardless of primer T_m.

All PCR products were purified using peq lab PCR purification kit (peq lab) according to manufacturer's protocol. DNA from each reaction was eluted with 30μl of sterilized ddH_2O.

2.1.5 Restriction enzyme digestion

In all cases when double restriction enzyme digests were performed, buffer conditions compatible with both enzymes were chosen (according to NEB restriction enzyme buffer chart). All digestions were carried out at 37^oC for 2.5 hours. Typical reaction mixture is shown in Table 2.3.

Reagent	Volumn (μl)
dH_2O	5
Reaction Buffer (10x)	2
BSA (20x)*	1
DNA (up to 1 μg)	10
Restriction enzyme 1 (10U/μl)	1
Restriction enzyme 2 (10U/μl)	1
Total	20

Table 2.3 Typical restriction enzyme digest for PCR products and vectors. *BSA was used according to restriction enzymes chosen.

Double digested DNA products were purified by electrophoresis using 1% argose gel (peq Lab) running in TAE buffer (40mM Tris Acetate 1mM EDTA pH 8.0). Gels were stained with 0.5 μg/ml ethidium bromide (Fisher) and visualized on UV-transilluminator. The desired vector or PCR product was excised from gel using sterile scalpel blade and purified using Gel Extraction Kit (pEQ Lab). Each reaction was eluted with 30 μl H_2O.

2.1.6 Ligation reactions

Ligation reactions were prepared as in Table 2.4 and incubated for 15 mins at room temperature using DNA ligation kit (TaKaRa Bio).

Reagent	Volumn (μl)
Digested DNA insert	7
Digested vector	1
TaKaRa ligation mix	8
Total	16

Table 2.4 Typical ligation reaction. An excess amount of insert DNA is added to increase the chance of ligation.

2.1.7 Transformation

All transformations of bacterial cells were carried out according to the manufactures' instructions supplied with competent cells (Bioline). A typical transformation was performed as follows: 1 μl of DNA (pre-incubated on ice, or 5ul DNA from ligation) was added to 50 μl of competent cells, incubated on ice for 30 mins, heat-shocked at 42^{o}C for 45 seconds, incubated on ice for 2 mins, incubation at 37^{o}C shaker for 60mins after addition of 900μl of LB and plated on appropriate media. All plates were incubated overnight at 37^{o}C.

2.1.8 Protein expression

All expression conditions were optimised, *E.coli* BL21(DE3) competent cells (Bioline) was selected, which were transformed according the manufacturer's protocol. One colony from an overnight plate was used to inoculate 5ml of LB media containing the appropriate antibiotics. The culture was grown overnight at 37 ^{o}C and used as starter

culture. 5ml of starter culture was used to inoculate 1L of LB media with appropriate antibiotics. The culture was grown at 37 ^{o}C and induced with 1mM IPTG when optical density (OD_{600}) reaches 0.6-0.8. Cells were harvested by centrifugation 4hours (2 hours for OutD-Cter) post induction at 5,000g for 20minutes.

2.2 Protein purification techniques

2.2.1 Bacterial lysis

All bacteria lysis was done at 4 ^{o}C. Bacteria pellets were fully resuspended in 10 ml/g (lysis buffer /pellet) lysis buffer according to individual purification protocol and supplemented with 30kU/g lysozyme (Novagen) to aid lysis and reduce lysate viscosity. Cells were lysed by sonication (VibraCell) on ice at 60% of power with 6 bursts of 15s and 30s cooling in between. The soluble protein fraction was obtained by centrifugation at 15,000g for 15 mins. Samples of lysed cells, soluble protein fraction and insoluble pellet were used for SDS-PAGE analysis.

2.2.2 His tag purification

All proteins were expressed as hexahistidine tagged fusion proteins and were purified as follows. Firstly, the cell pellet was lysed (as above) in Ni column binding buffer (20mM Tris, 400mM NaCl, 20mM imidazole pH8.0). The soluble protein fraction was then separated by centrifugation at 15,000g for 15 mins and applied to the pre-packed His-Trap column (GE Healthcare). The colum is pre-equilibrated in Ni-column binding buffer, using a flow rate of 1ml/min. The non-specifically binding proteins were eliminated by washing with 20 column volumes (CV) of Ni column washing buffer (20mM Tris, 400mM NaCl, 20mM imidazole pH 6.5) until no more protein was in the flow through (tested using Bradford solution).

If uncleaved protein was required, the target was eluted using eluting buffer (20mM Tris, 400mM NaCl, 500mM imidazole, pH6.5) again until no further protein came through (tested using Bradford solution). The eluted fractions were checked using SDS-PAGE.

Alternatively, if the His tag was required to be cleaved off, the column was equilibrated with thrombin cleavage buffer (20mM Tris, 150mMNaCl, 2.5mM $CaCl_2$ pH8.4). Then 10U/ml of thrombin in thrombin cleavage buffer was loaded onto the column and incubated at 4^oC overnight. The target protein was then eluted with 20mM Tris pH8.0.

2.2.3 GST tag purification

In the case of OutC-HR and OutD-Cter, pGEX-6P-3 was used to produce N-terminal cleavable GST tagged protein. To purify the protein using the glutathione column, the cell pellet was resuspended and lysed (as described in section 2.2.1) in 20mM PBS pH7.4. The soluble protein fraction was then applied to a bench-run column with 5ml glutathione Sepharose resin (GE Healthcare) pre-equlibrated with PBS. The loaded glutathione Sepharose resin was washed with 20 CV PBS till no further protein was washed off (tested by Bradford solution).

If the protein was to be cleaved once eluted from the resin, the uncleaved protein was eluted with 20mM reduced glutathione in PBS pH7.4, until no protein come through (tested by Bradford solution). The eluate was then dialyzed against the PreScission Protease cleaving buffer. 20U PreScission Protease (GE Healthcare) was then added to the elutate and incubated at 4^OC for 16 hours. At the end of this incubation the target protein is in a mixture with equal, if not greater, molar amount of GST tag and likely to

have more contamination than that produced by cleavage from the resin.

If the protein was cleaved on resin, buffer exchange with PreScission Protease (GE Healthcare) cleaving buffer (50mM Tris, 150mM NaCl, 1mM EDTA, 1mM DTT pH7.0) preceded the addition of 5ml of PreScission Protease cleaving buffer with 20U of PreScission Protease and incubation at $4^{O}C$ for 16 hours. The target protein was then eluted with PreScission Protease buffer in 2 ml fractions. Aliquot of each elute fraction was checked using SDS-PAGE.

2.2.4 Purifying protein from the periplasm

OutS protein was produced with a PelB signal sequence, which targets the protein to the periplasm (Fries, Ihrig et al. 2007). To purify the protein from the periplasm, the freshly harvested cell pellet from the expression cultures were resuspended in ice-cold 30 mM Tris–HCl pH 8.0 with 1 mM EDTA and the periplasmic protein was extracted from the periplasm by addition of the same volume of ice-cold 40% w/v sucrose and incubation on ice for 10 min. The soluble OutS was then separated by centrifugation at 15,000g for 15 mins. Subsequently, the extracted protein solution was dialysed against 20 mM Tris 150mM NaCl (pH 7.0) and purified using size exclusion chromatography.

2.2.5 Size exclusion chromatography

The eluate from previous method was concentrated to 1-2ml and loaded on to a Superdex prep grade S75/200 HR 10/30 column (GE Healthcare) using ÄKTA purifier system (GE Healthcare). The column was pre-equilibrated with buffer and run at 0.5ml/min with 1 ml fractions collected. Fractions were checked using SDS-PAGE and appropriate peaks were pooled, concentrated and stored at $-80^{o}C$.

2.2.6 SDS-PAGE

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE), is a technique used to separate proteins according to their electrophoretic mobility, a function of the length of a polypeptide chain and its charge. It was used to track the expression and purification of proteins.

According to the size of the protein to be resolved on the gel, 12% and 18% polyacrylamide gels were used. The composition of the gels is given in Table 2.5. The gels are made by adding TEMED just before pouring the gel mixture into the Bio-Rad Gel cast (Bio-Rad). The resolving gel was poured into the cast first and followed by the stacking gel and a comb for 10 or 15 wells was inserted and before the gel was left to set.

Solution (ml)	4% Stacking (1.5ml)	12% Resolving (5ml)	18% Resolving (5ml)
H_2O	1.02	1.65	1.3
30% acrylamide	0.255	2	3
1.5M Tris (pH8.8)	-	1.25	1.25
1.0M Tris (pH6.8)	0.1875	-	-
10% SDS	0.015	0.05	0.05
10% ammonium persulfate	0.015	0.05	0.05
TEMED	0.0015	0.002	0.002

Table 2.1 Recipe for SDS-PAGE. Resolving gel are used at 12% or 18% and stacking gel is 4% for either gels.

Protein samples were prepared by adding equal amount of 2 x SDS protein loading buffer (100mM Tris, 200mM DTT, 4% SDS, 0.2% bromophenol blue, 20% glycerol) and protein sample and heated at 100 ^{o}C for 5 mins. Gels were run in SDS Tris-Glycine buffer (25mM Tris, 192mM Glycine, 0.1% SDS, pH 8.3) at 150V till the protein loading buffer dye ran out in a Bio-Rad Mini-Gel cell on a Bio-Rad powerPac 300 (Bio-Rad). The gel was developed by staining in comassie blue dye (0.1% comassie blue,

30% methanol, 10% acetic acid) for 10 minutes on shaker and destained in destain buffer (10% acetic acid, 30% methanol) until the protein bands were clearly observed.

2.2.7 NMR sample preparation

Recombinant uniformly ^{15}N- ^{13}C- labeled full length OutC-HRF3 (or OutD-N0) was expressed using the pET-14b vector (Novagen) in BL21(DE3) (Novagen), grown in M9 minimal media. 1L M9 minimal media contained Na_2HPO_4 (6g), KH_2PO_4 (3g), NaCl (o.5g), $MgS0_4$ (0.25g), $CaCl_2$ (0.015g), $FeS0_4$ (0.015g), thiamine (0.001g), Biotin (0.001g), $^{15}NH_4Cl$ (1g) (99% ^{15}N from Cambridge Isotope laboratories Inc.), ^{13}C D-glucose (2g) ($^{13}C6$, 99% from Cambridge Isotope laboratories Inc.) and ampicillin (100mg). All minimal media was 0.2μm filter sterilized and 15ml induced expression cultures were grown at 37^{o}C in LB as above.

NMR samples contained 0.5mM (or as stated) uniformly labeled OutC- HRF3(or OutD-N0) in 90% H_2O/ 10% D_2O containing 20mM Tris pH 7.0 . NMR samples were initially optimised against different ionic strength and pH . Protein solubility and homogeneity did not shown much difference againest different conditions, whereas neutral and alkaline buffer conditions were prefered. Therefore a no salt buffer was choosen for achieving lower signal to noise ratio form NMR spectroscopy. pH 7.0 was choosen as higher pH will speed up amide proton exchange rate reduce spectra resolution.

2.2.8 Protein concentration determination

Protein concentration was determined directly by spectroscopic analysis using UV-Vis spectrophotometer (HITACHI). Protein absorbance at 280nm was measured. Concentration was determined according to:

$$A_{280} = \varepsilon \, . \, c \, . \, l$$

Where c is the molar concentration, A_{280} is the absorbance at 280nm, ε is the molar extinction coefficient (calculated using ExPASy ProtPrama) and l is the path-length of the cuvette.

The protein A_{260}/A_{280} was measured to check DNA contaminations. A_{260}/A_{280} below 0.7 was considered to have no DNA contamination.

2.3 Biochemical and biophysical techniques

2.3.1 Limited proteolysis

Sequencing grade modified trypsin (Sigma) was used in all limited proteolysis experiment at a ratio to target protein of 1/1000, 1/5000, 1/10,000, 1/50,000, 1/100,000. All reactions were at room temperature in trypsin buffer (50mM Tris, 20mM $CaCl_2$, pH 8.0). The trypsin digest was followed over time by taking aliquots at various time intervals, terminating the reaction with 0.5mM phenylmethylsulfonyl fluoride, and assessing the effects by SDS-PAGE. The solubility of the digested products was also checked using a native gel.

2.3.2 Mass spectrometry

The trypsin digested fragments mass were detected on mass spectroscopy with the help of Dr Robin Maytum. To prepare sample, formic acid was mixed with the protein sample (~10ug/ml) to a final concentration of 1% (v/v). The sample was first bound to a C5 reverse-phase HPLC column and then eluted with an acetonitrile gradient. The eluted fractions from the HPLC column were assessed by mass spectroscopy.

2.3.3 N-terminal sequencing

N-terminal sequencing of the OutC-HR fragments was by Protein & Nucleic Acid Chemistry Facility (University of Cambridge).

2.3.4 Dynamic light scattering

Dynamic light scattering is a technique used to determine the size distribution profile of small particles in suspension. In dynamic light scattering, the speed at which the particles are diffusing due to Brownian motion is measured. This is done by measuring the rate at which the intensity of the scattered light fluctuates when detected using a suitable optical arrangement. The rate of the scattering intensity fluctuation occurring depends on the size of particles. The small particles cause the intensity to fluctuate more rapidly than the large ones. The size of a particle is calculated from diffusion coefficientusing the Stokes-Einstein equiation:

$$D=k_BT/6\pi\eta r$$

where r is the radius of the protein, k_B is the Boltzmann constant, T is the temperature in Kelvin degrees and η is the viscosity of the solvent. In this study protein molecular size is caculated from the radius using raleigh sphere module.

Dynamic light scattering was measured using a DynaPro Molecular sizing instrument running DYNAMICS V6 software. The quartz cuvettes (45 μl) were prewashed with 1% Triton and subsequently with water and dried with compressed nitrogen air. The exterior surface of cuvette was wiped with ethanol and lens tissue to remove any dirt on the outside surface. The protein concentration used was 4mg/ml (or as stated). Each sample was passed through a 0.2μm filter and a minimal 20 measurements were recorded at 20^{o}C.

2.3.5 Circular dichroism

Circular Dichroism (CD) is an excellent method to give an indication of the secondary structure of the protein. Although it is not possible to give detailed residue-specific information as obtainable from NMR and crystallography, it can provide secondary structure information using only small amounts of protein and it is extremely sensitive to changes in secondary structure and can also be used to monitor the conformational changes (for instance as a function of temperature, pH, or ionic strength).

CD measures differential absorption of left- and right- handed circular polarized light. The far UV CD bands of protein reflect the secondary structure, such as α-helix, β-sheet and unordered content (Figure. 2.1).

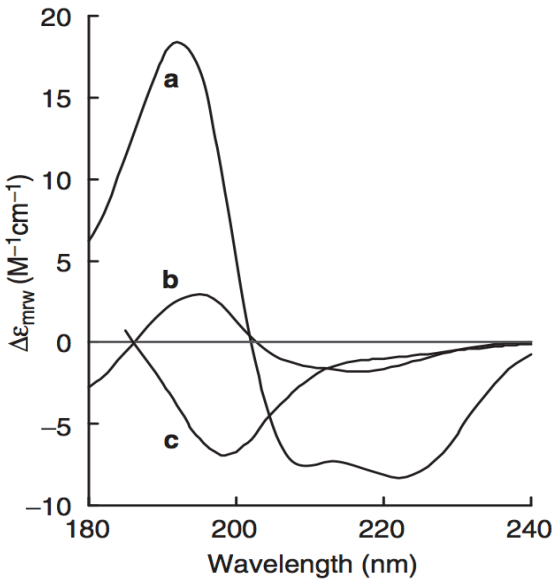


Figure 2.1 Typical far-UV CD spectra of (a) myoglobin (all-α: 4mbn.pdb); (b) prealbumin (all-β: 2pab.pdb), and (c) acid denatured staphylococcal nuclease at pH 6.2 and 6 _C (irregular). Adapted from (Martin and Schilstra 2008).

All samples prepared for CD were purified to the highest purity and checked at 260/280 absorbance for DNA contamination. The final protein concentration was determined using absorbance at 280 nm in 6M guanidine hydrochloride (Martin and Schilstra

2008). Protein samples used for CD were at a concentration of 0.05mg/ml in 20mM Tris 150mM NaCl (pH 7.0).

Far-UV CD measurements were made using a Jasco J-715 spectropolarimeter equipped with a PTC-348WI temperature controller. Spectra were recorded in 20mM Tris, 150mM NaCl (pH7.0) at 15 °C using 1mm path length fused silica cuvettes, The spectra are presented as differential absorbance after baseline subtraction.

2.3.6 Fluorescence spectroscopy

Fluorescence spectroscopy is a type of electromagnetic spectroscopy which analyzes fluorescence from a sample of light, usually ultraviolet light, that excites the electrons in molecules of certain compounds and causes them to emit light of a lower energy. Aromatic residues are good natural fluorescence probes for protein studies. For tryptophan, the emission spectra usually peak between 310-350nm, depending on the local environment, the more exposed tryptophan tend to peak more towards 350nm and have higher signal.

Fluorescence data were collected using a Jasco FP-6300 Spectrefluoremeter. To avoid exciting tyrosyl side chains, an excitation wavelength of 295nm was used. Emission spectra were recorded at 15 °C in steps of 2 nm from 310 to 400 nn. The fluorescence signal at 340 nm was plotted to calculate K_d. Samples are prepared in 20mM Tris, 150mM NaCl, pH7.0. OutS spectra was measured at 1μM in buffer, 50μM of OutD691-708 prepared in 1μM OutS was titrated into OutS solution.

2.3.7 Preparation of OutC and OutD coexpression pull down assays

The cell pellet from 1L of culture was resuspended in 30 ml of Tris-Fix buffer (50 mM Tris, 150 mM NaCl, 1 mM EDTA pH 7.5) and sonicated. The soluble and insoluble protein fractions were separated by centrifugation at 15,000 g for 15 mins. The GST-$OutC^{77-162}$ produced appeared to be insoluble. Therefore, the insoluble protein pellet was then resuspended in 5 ml of 8 M urea and topped up with 20 ml of Tris-Fix buffer gradually with mixing. The urea refolded soluble protein is separated from the insoluble protein by centrifugation at 15,000 g for 15 min.

Subsequently, the urea refolded soluble protein was loaded onto 5 ml pre-equlibrated Ni chelating beads (GE healthcare) and incubated at room temperature with mixing for 20 mins. Non-specific binding proteins were washed off using 50 ml of Washing Buffer (50mM Tris, 400mM NaCl, 20mM imidazole pH6.5). Protein was eluted using 5 ml of Eluting Buffer (50mM Tris, 400mM NaCl, 20mM imdiazole pH6.5).

2.3.8 Thermofluor

Thermofluor is a high-throughput protein stability test. In this case, protein OutC-HRF3 and OutD-N0 are used to test the stability of the proteins in different buffer conditions. The buffer that gives the higher melting point tends to provide the protein with better stability.

Protein folding/ unfolding was monitored by the dye SYPRO-Orange (Invitrogen). The SYPRO-orange signal is highly quenched in an aqueous environment. As the protein unfolds, hydrophobic surfaces that are buried in the native protein become exposed to solvent and Sypro-orange binds to these hydrophobic sites resulting a signal.

Thermofluor analysis exploited a qPCR machine with 96 well PCR detection plate (AB gene AB-1100). The screened buffer conditions are shown in Figure 2.6. For each screen mixture a 25µl of screen mix was prepared, including 12.5µl screening buffer, 5µl 25xSypro orange (Invitrogen, prepared by mixing 5µl of 5000x stock with 955µl of water) and 7.5µl protein (starting with 5mg/ml and diluting if necessary). The screen was between 25-95 ^{O}C with one degree intervals. For each temperature, a 45 second incubation time was given between measurements. The emission signal was measured at 525nm.

A

		1	**2**	**3**	**4**	**5**	**6**
		100mM Na-Acetate pH 4.5	100mM Bis-Tris pH 5.5	100mM Na-Citrate pH 6.5	100mM HEPES pH 7.5	100mM Tris/HCl pH 8.5	100mM CHES pH 9.5
A	0mM NaCl						
B	150mM NaCl						
C	500mM NaCl						
D	150mM NaCl	10% Glycerol	10% Glycerol	10% Glycerol	10% Glycerol	10% Glycerol	10% Glycerol
E	150mM NaCl	5mM B-OG	5mM B-OG	5mM B-OG	5mM B-OG	5mM B-OG	5mM B-OG
F	150mM NaCl	50mM Glycine	50mM Glycine	50mM Glycine	50mM Glycine	50mM Glycine	50mM Glycine
G	150mM NaCl	50mM Arg+Glu	50mM Arg+Glu	50mM Arg+Glu	50mM Arg+Glu	50mM Arg+Glu	50mM Arg+Glu
H	150mM NaCl	100mM Urea	100mM Urea	100mM Urea	100mM Urea	100mM Urea	100mM Urea

B

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
A	1	2	3	4	5	6	7	8	9	10	11	12
B	13	14	15	16	17	18	19	20	21	22	23	24
C	25	26	27	28	29	30	31	32	33	34	35	36
D	37	38	39	40	41	42	43	44	45	46	47	48
E	49	50	51	52	53	54	55	56	57	58	59	60
F	61	62	63	64	65	66	67	68	69	70	71	72
G	73	74	75	76	77	78	79	80	81	82	83	84
H	85	86	87	88	89	90	91	92	93	94	95	96

Table 2.2 Conditions used in 96 well thermofluor buffer screen. (A) Buffer conditions used in the screen. (B) Numbering system of the 96 well plates.

Chapter 3 NMR Spectroscopy: Materials and Methods

3.1 Overview

There are two techniques used routinely to determine protein structure: X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy. X-ray crystallography is the direct measurement of the diffraction from the protein crystal to calculate, subject to solving the phase problem, the electron density and from that the protein structure. In contrast NMR spectroscopy uses a number of techniques including establishing the chemical shifts measured for each nuclear magnetic active atom and measuring proton-proton distances to determine the protein structure.

Both X-ray crystallography and NMR spectroscopy have their limitations. X –ray crystallography is dependent on the production of diffracting protein crystals, which is rate limiting as it can involve extensive crystallization screens and optimization of conditions. Structure determination by NMR is limited by the size of the protein. As the size of the protein of interest gets bigger, the difficulty involved in solving the structure increases dramatically.

NMR is also a very sensitive method to measure the time dependent changes of a protein. Therefore, it offers a great way to explore protein/ligand interactions and protein dynamics.

3. 2 NMR spectroscopy

3.2.1 Spectroscopy

NMR spectra were acquired at 15^0C on Varian Inova 800 and 600 and Bruker Avance 700 and 600 MHz spectrometers, all equipped for $^{1}H/^{15}N/^{13}C$ triple-resonance

experiments. The Bruker spectrometers were controlled by TOPSPIN software (http://www.bruker-biospin.com/topspin3.html) and the Varian spectrometers by VNMR software (http://www.varianinc.com/cgi-bin/nav?products/nmr/software/).

All NMR spectra were processed using NMRPipe (Delaglio, Grzesiek et al. 1995) and analyzed using Computer-Aided Resonance Assignment (CARA) version 1.5.5/NEASY (http://www.nmr.ch/).

^{15}N singly or ^{13}C- and^{15}N-doubly isotropically labeled protein were used to record various spectrum. Protein used to record 3D experiments was around 0.5mM. Sample concentrations ranging from 25μM to 0.5mM were used to record 2D spectra depending on the nature of the experiment and the time available.

^{1}H-^{15}N Heternuclear Single Quantum Coherence (HSQC) spectra and ^{1}H-^{15}N Sofast-HMQC were used for purposes such as sample evaluation, thermal melt analysis, titration experiments and for assignment.

Various 3D experiments (shown in Table 3.1) were used for backbone assignment, side-chain assignment and to generate distance restraints for OutC-HRF3. Backbone assignment was done for OutD-N0 and OutD-Cter. The buffer conditions used for OutC-HRF3 and OutD-N0 was 20mM Tris pH7.0 and 10% D_2O. The buffer conditions used for OutD-Cter was 20mM Tris, 150mM NaCl pH7.0 and 10% D_2O.

Experiment	Purpose of experiment
1H NMR	Sample evaluation Reference spectrum Titration
^{1}H-^{15}N HSQC	
^{1}H-^{15}N HMQC	
HNCACB	Backbone assignment
CBCA(CO)NH	
HNCO	
HNCA	
CBCAHA	Side-chain assignment
HACA(CO)N	
HCCH-TOCSY	
^{1}H-^{15}N NOESY	Distance restraints
^{1}H-^{13}C NOESY	

Table 3.1 Summary of the types of NMR spectra recorded during this work.

3.2.2 Spectral analysis and resonance assignment

Standard triple resonance backbone assignment of protein using HNCACB and CBCACONH spectra was used (Figure 3.1). The idea is that the HNCACB spectra correlates each NH group with the chemical shift of the current residue's CA and CB (strongly) and that of neighboring residue's CA and CBs (weakly). The CBCA(CO)NH spectra only correlates the NH group of the current residue to the preceding CA and CB chemical shifts. By comparing both these spectra, the current residue and its preceding CA CB resonances can be identified. The succeeding amide proton chemical shift was identified using the chemical shifts of the preceding CA CB to search for the amide giving same chemical shift within the current residue's CA, CB on HNCACB spectra. For proline residues, where amide proton is absent, CBCAHA and HACA(CO)N were used to help with CA CB and HA identification. Assignment of side-chain resonances was achieved by using HCCH-TOCSY, ^{1}H-^{13}C NOESY and ^{1}H-^{15}N NOESY spectra.

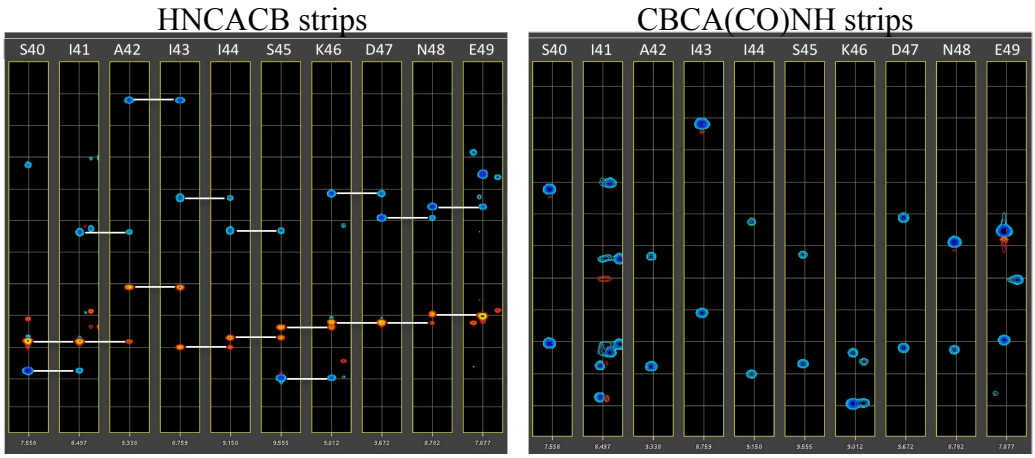


Figure 3.1 Resonance assignment using HNCACB and HN(CA)CO spectra.

3.3 Dynamic studies by NMR

NMR measured chemical shift depends on the local magnetic field. If there is molecular motion, then there is different shielding at different time points which creates different chemical shifts. By using this property, NMR can be used to study protein dynamics.

NMR is sensitive to rather slow molecular motions. NMR relaxation studies can therefore provide detailed information pertaining to the internal dynamics occurring in proteins. It is well established that despite the close packing in globular proteins, there are substantial fluctuations on the picosecond time scale. Different physical processes are responsible for the relaxation of the components of the nuclear spin magnetization vector M parallel and perpendicular to the externally applied magnetic field B_0. These two principle relaxation processes are termed longitudinal (or T_1, spin-lattice) relaxation and transverse (or T_2, spin-spin) relaxation. The measurement of ^{15}N relaxation rate is particularly useful for obtaining dynamic information, as the relaxation of these nuclei is governed predominantly by the dipolar interaction with directly bound protons. Therefore by measurement of ^{15}N relaxation parameters, the hydrodynamic properties of protein can be gained.

^{15}N T_1, T_2 and ^{1}H-^{15}N NOE enhancement relaxation experiments were performed at 600MHz (at 15^{o}C) using standard acquisition methods (Kay LE 1989).

3.3.1 Longitudinal and transverse relaxation

T_1 and T_2 values are sensitive to different motional frequencies. T_1 relaxation times provide motional properties with a frequency of approximately 10^{8}-$10^{12}s^{-1}$, while T_2 values, in addition to depending on motions occurring at these high frequencies, are also sensitive to dynamics on the micro and millisecond time scale. By obtaining both T_1 and T_2 relaxation values, it is possible to obtain dynamic motion information over a large frequency range.

For small molecules or unstructured region of proteins, which reorient rapidly in solution, the T_1 and T_2 relaxation times are similar (Figure 3.2). As the protein becomes larger and more structured, it tumbles more slowly in solution.

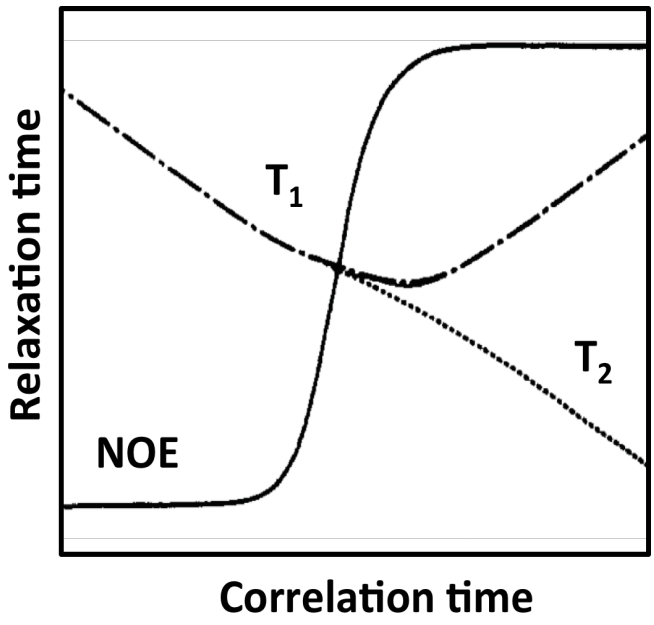


Figure 3.2 Schematic representation of T_1, T_2 and NOE relaxation times as a function of correlation time (Kay LE 1989).

The ratio between the average T_1 and T_2 relaxation values for a given molecule can be used to calculate its rotational correlation time (τ_c) as follows:

$$\tau_c=(4\pi\nu_N)\,(6T_1/T_2\,-7)^{1/2}$$

Where ν_N is the atom resonance frequency in Hz, the ^{15}N resonance frequency is 60,000,000Hz when measured at 600MHz spectroscopy.

τ_c is the time it takes the particle to rotate by one radian. It depends on the particle size. As a general rule of thumb, the τ_c of a monomeric protein in solution in nanoseconds is approximately 0.6 times its molecular weight in kDa. Moreover, when calculated correlation times are similar throughout the protein backbone this suggests isotropic motion of the protein molecule.

3.3.2 The Nuclear Overhauser Effect (NOE)

The Nuclear Overhauser Effect (NOE) is a phenomenon observed by cross relaxation between dipolar coupled spin systems. It is exceptionally important tool for structural studies.

For dynamic studies, transient NOE experiments were used. ^{15}N NOE is 5 to 30 times more sensitive to internal dynamics than T_1 and T_2. Residues located on loops can show shorter NOE than any of the non-terminal residues. Intense negative NOEs suggest contributions from rapid internal motions and can usually be observed for the terminal residues. In multidomain proteins, the backbone correlation time can also give some assessment of the degree of intra-domain association, such as for the same protein size, domains arranged in a "beads on a string" manner gives shorter correlation time than those of a single domain or domains with strong inter-domain contacts. If the dipolar

interaction is considered exclusively, for rigid proteins tumbling isotropically the maximum NOE possible is +0.82 (Kay LE 1989).

For structure calculations, NOE restraints were obtained form 3D ^{13}C- and ^{15}N-edited NOESY-HSQC spectra. From these experiments, homonuclear ^{1}H-^{1}H NOEs can be clearly assigned even in overcrowded regions. The intensities of the peaks reflect the distance between two protons. Based on the available chemical shift assignments, the NOE data are essential for the secondary and tertiary structure calculations.

3.4 Secondary structure prediction by TALOS

The initial protein NMR structural characterization began with prediction of the secondary structure. It is well established that when a polypeptide chain forms secondary structure, the amino acid residues display a characterized chemical shift change in relation to their standard random coil values (Wishart and Sykes 1994). Namely, the chemical shifts of CA nuclei experience an up-field shift when present in β-strand structure and a down-field shift when in α-helices.

Protein backbone chemical shifts are exquisitely sensitive to local conformation, such as bond angles. The protein backbone dihedral angels (ϕ and ψ , Figure 3.3) are characteristic to the secondary protein structure, this correlation is well presented as Ramachandran plot (Ramachandran, Ramakrishnan et al. 1963). By using the inverse of this relation, TALOS+ searches the input protein sequence in a data base of 200 proteins for the 10 best matches to the secondary chemical shits of given residual in a target protein along with its two flanking neighbours (a residual triplet). If there is a consensus ϕ and ψ angle among the 10 best database matches, then database triplet structures are used to predict the secondary structure of the target residues. TALOS incorporates the

assignments of ^{15}N, ^{1}H, $^{13}C\alpha$, $^{13}C\beta$, $H\alpha$, $H\beta$ and ^{13}C chemical shifts of each amino acid for secondary structural prediction (Figure 3.4).

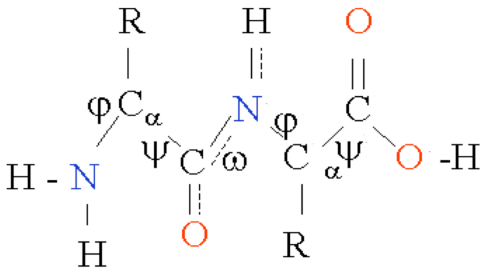


Figure 3.3 Protein dihedral angels. Peptide bond (ω) phi (ϕ) and psi (ψ)torsion angles are shown.

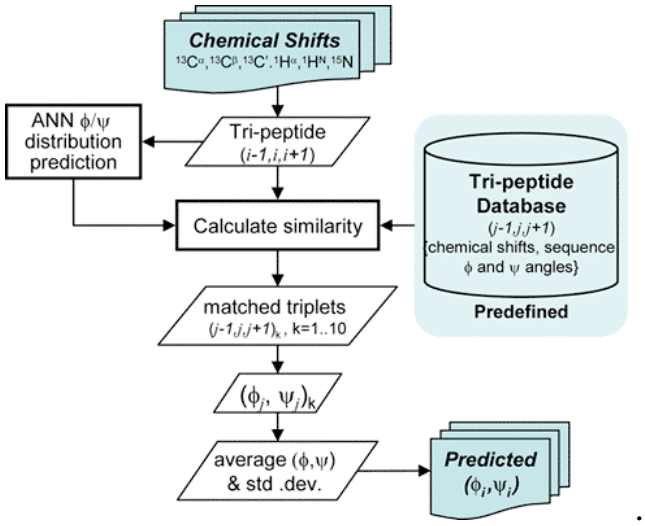


Figure 3.4 Schematic representation of the calculations by the TALOS+ program (Shen Y 2009)

3.5 Residual dipolar coupling

Residual dipolar coupling (RDC) is an NMR parameter that provides a measure of dipolar interaction between a pair of atoms, such as ^{1}H-^{15}N (Bax and Grishaev 2005). In solution, RDCs are averaged to zero as they are very sensitive functions of the time-averaged orientation of the corresponding internuclear vectors and thereby offer highly precise structural information. RDC can be measured in partially aligned protein samples, which can be achieved using a diluted liquid crystalline phase. The alignment

happens in very non-invasive manner so that the resulting residual dipolar coupling is about 1000 times smaller than the original dipolar coupling interaction, providing orientation information without affecting the shape and dynamics of the protein.

OutC-HRF3 anisotropic liquid crystalline state was achieved according to Rückert and Otting's method (Rückert and Otting 2000). Lα phase crystals were prepared by dissolving C5E8 in 90% 20mM Tris (pH7.0)/10%D_2O, and adding n-octanol in microlitre steps to the desired final molar ratio under vigorous shaking. The solution was biphasic at low alcohol concentrations and became instantaneously transparent and opalescent upon crossing the Lα phase boundary. The Lα phase is relatively viscous immobilizing any small air bubbles presence in solution. Equal volumes of 0.5mM ^{15}N labeled OutC-HRF3 were mixed with the Lα phase solution to achieve the final anisotropic sample. The presence of an ordered lamellar phase was monitored by observation of quadrupolar splitting of the ^{2}H NMR signal of the solvent.

3.6 ARIA 1.2

ARIA 1.2 (Ambiguous Restraints for Iterative Assignment) is the program used for OutC-HRF3 structure calculation (Linge, Habeck et al. 2003). Due to the high complexity of the assignment problem, most of the NOEs cannot be directly converted into an unambiguous inter-proton distance. ARIA takes structural information including, torsion angles, hydrogen bonds, residual dipolar coupling and identified unambiguous NOE list together with ^{1}H-^{15}N-NOESY and ^{1}H-^{13}C-NOESY peak list and calibrated peak intensity. The program uses the input data to automatically assign NOE peaks and calculate the structure. The parameters used for structure simulation are shown in Appendix 2.

3.7 NMR titration

NMR titration is achieved by recording a ^{1}H-^{15}N HSQC/ ^{1}H-^{15}N Sofast-HMQC spectra of the labeled protein. In serial titrations, cross titrations were performed to eliminate the dilution of the labeled protein. In cross titration, two samples were prepared; one sample was labeled protein only and the second sample had the same concentration of labeled protein and a high equivalent (i.e. 10 equivalents) of unlabeled titration partner. During the titraion the concentration of labeled protein did not change, so it was not necessary to correct for dilution.

3.8 HADDOCK

The program HADDOCK is a protein-protein docking algorithm which uses biochemical and biophysical information to provide a model of the protein-protein complex (Dominguez, Boelens et al. 2003). It uses experimental data (i.e. chemical shift perturbation data, site specific mutagenesis data). The information on interacting residues is introduced as ambiguous interaction restrains (AIRs) to drive the docking. In the case of NMR titration data, the active residues correspond to all residues showing a significant chemical shift perturbation upon complex formation as well as high solvent accessibility in the free form of protein. The passive residues correspond to the residues that show a less significant chemical shift perturbation and/or that are surface neighbors of the active residues with high solvent accessibility (>50%). The AIRs files are generated by entering active and passive residue on http://www.nmr.chem.uu.nl/haddock/. 1000 docking structures were generated for each calculation. The output structures are ranked according to their intermolecular energy. The best 100 structures were used for the water structure refinement.

Chapter 4: Solution structure of OutC-HRF3

4.1 Overview

In order to understand the function of the GspC homology region at the molecular level, the structure of OutC-HRF3 (104 residues, 11.1kDa, sequence shown in Figure 4.2 and Appendix1) was solved by NMR spectroscopy. A combination of manual and automatic methods were employed for the structure determination. Manual structure calculations were performed using constraints derived from ^{1}H-^{15}N and ^{1}H-^{13}C NOESY spectra, hydrogen bonds that were detected from the direct HNCO measurement, residual dipolar coupling measurements on a weakly aligned liquid crystalline media and backbone torsion angle restrains derived from TALOS. ARIA 1.2 was used for automatic structural calculation using unassigned NOE peaks, chemical shifts and hydrogen bonds. Residual dipolar couplings were introduced into the final round of structure calculation. The CS-ROSETTA predicted structure is consistent with final structure ; the OutC-HRF3 is revealed to possess a β−sandwich like fold.

4.2 OutC- HRF3 structure identification

GspC is a biotropic protein. It has a short N terminus located in cytoplasm, a transmembrane segment and a long C-terminal region located in the periplasmic region. Within the periplasmic region, there are two domains involved in function: the homology region and a domain involved in protein protein interactions, which is in most cases a PDZ domain, or a coiled-coil domain.

The OutC-HR construct was provided by the colabrator Dr. Shevchik (University of Lyon); it consists of the periplasmic region of OutC, but excludes the PDZ domain. A series of cystallization trails were carried out with no crystal being obtained.

Limited proteolysis is a technique with a good track record of helping to identify folded domains amenable for structural studies. The folded segments or domains of a protein are more resistant to proteolysis than the loops between domains because potential cleavage sites are less accessiblein the folded domains.. This technique was applied to OutC-HR to identify the more structured part. The protease trypsin (Sigma) was used to challenge OutC-HR. There are three major fragments obeserved in the result of limited proteolysis (Figure 4.1A), OutC-HRF1, OutC-HRF2, and OutC-HRF3. As the order of OutC-HRF1 and OutC-HRF2 appeared; OutC-HRF1 shows up first and then OutC-HRF2 is starting to appear only when most of OutC-HR converted to OutC-HRF1; the OutC-HRF2 is a digested product from OutC-HRF1. Similar pattern also apply to OutC-HRF2 and OutC-HRF3, OutC-HRF3 is a further digested product of OutC-HRF2. Therefore, truncations happen in a temporal order, with the exposed sites cleaved first. From the temporal order of occurrence on the SDS-PAGE the corresponding fragment in the native gel was established. Native-PAGE was used to confirm that the proteolytic fragments obtained were soluble and not aggregating (Figure 4.1B).

The N-terminal protein sequences of the three proteolytic fragments were characterized by sequencing (Cambridge University sequencing service) and mass spectroscopy. The corresponding molecular masses and sequences are shown in Figure 4.2.

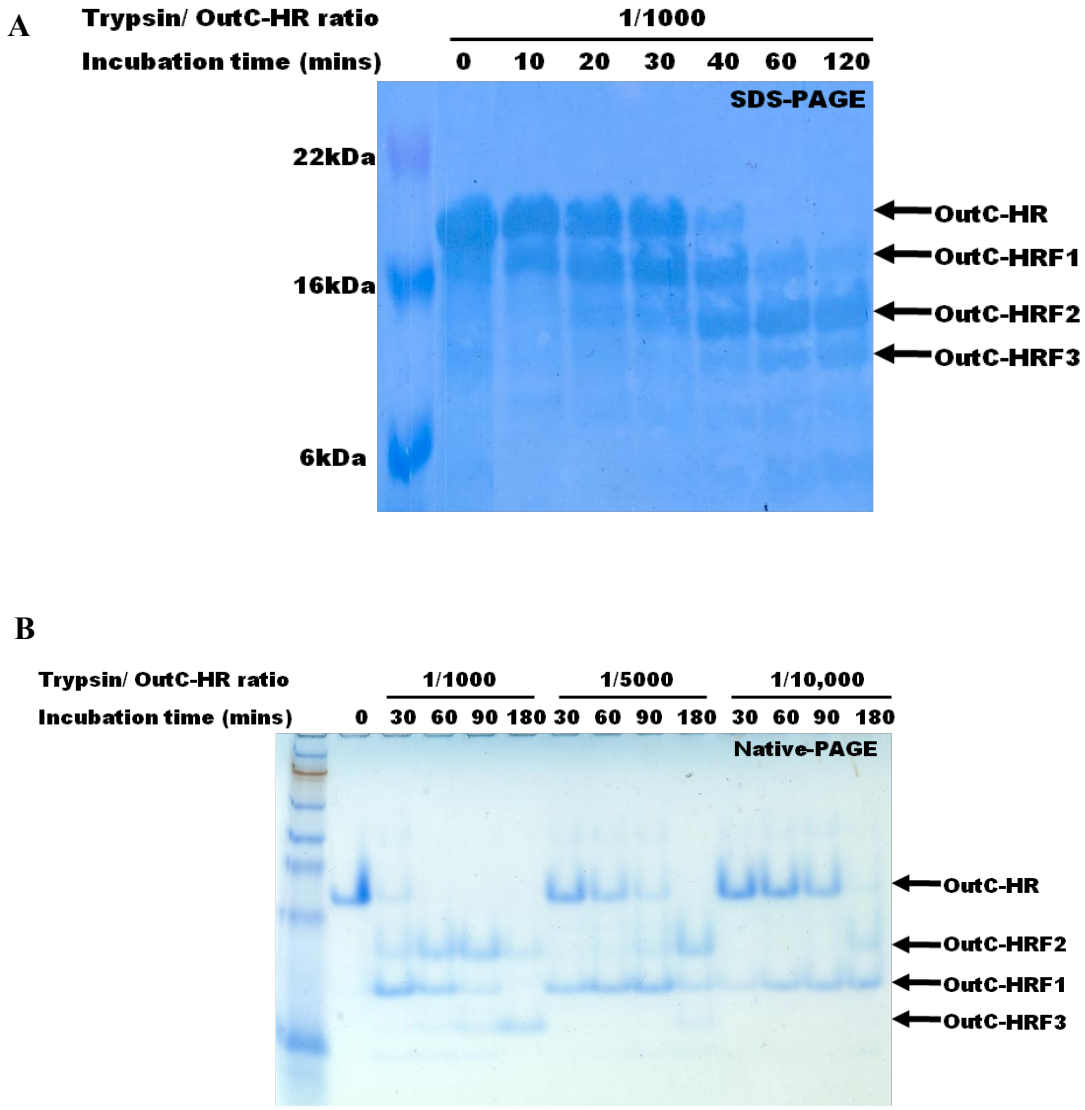
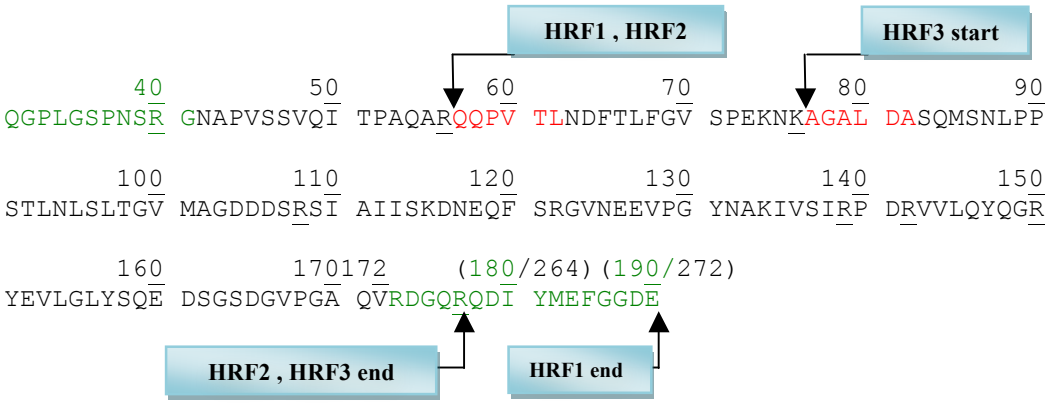


Figure 4.1 Limited proteolysis of OutC-HR. (A), SDS-PAGE of OutC-HR after limited proteolysis. (B), native PAGE of OutC-HR after limited proteolysis. The trypsin to Out-HR ratio ranges from 1/1000 to 1/100,000. The incubation time as indicated.

A

Fragment	N-terminal starts from	Mass detected from mass spec (Da)	Calculated mass (Da)	Region
OurC-HR	--	--	16773.4	42-190
OutC-HRF1	57	14330.5	14329.7	57-190
OutC-HRF2	57	12588.6	12587.9	57-177
OutC-HRF3	77	10344.2	10343.4	77-177

B



C

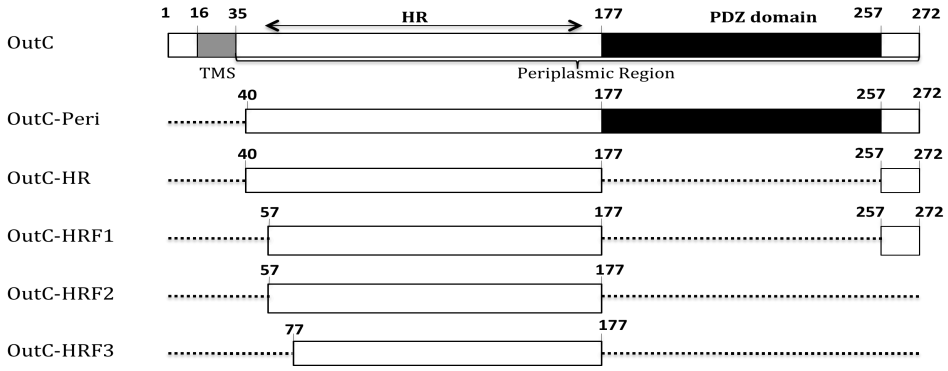


Figure 4.2 Sequencing results reveal the sequence of the putative domains. (A) Summary of the N-terminal sequencing and mass spectroscopy results. (B) OutC-HR construct sequence. Sequence identity of identified protein. OutC residues numbered in black, N-terminal sequencing detected residues in red, residues in OutC-HR in green. (C) Schematic diagram of OutC constructs.

4.3 OutC-HRF3 expression

4.3.1 Sub-cloning outC-hrf3

In the expectation that the smaller fragments corresponded to the more structured regions, OutC-HRF2 and OutC-HRF3 were chosen to be made recombinantly for further structural studies. Both corresponding sequences were cloned into pET-14b. The pET-14b vector carries a cleavable N-terminal His tag sequence, which facilitates protein purification. The ligation between outC-hrf3 insert and pET-14b vector was checked by *Xho1 Blp1* double digestion (Figure 4.3A). The nucleic acid sequence was checked by sequencing (MWG).

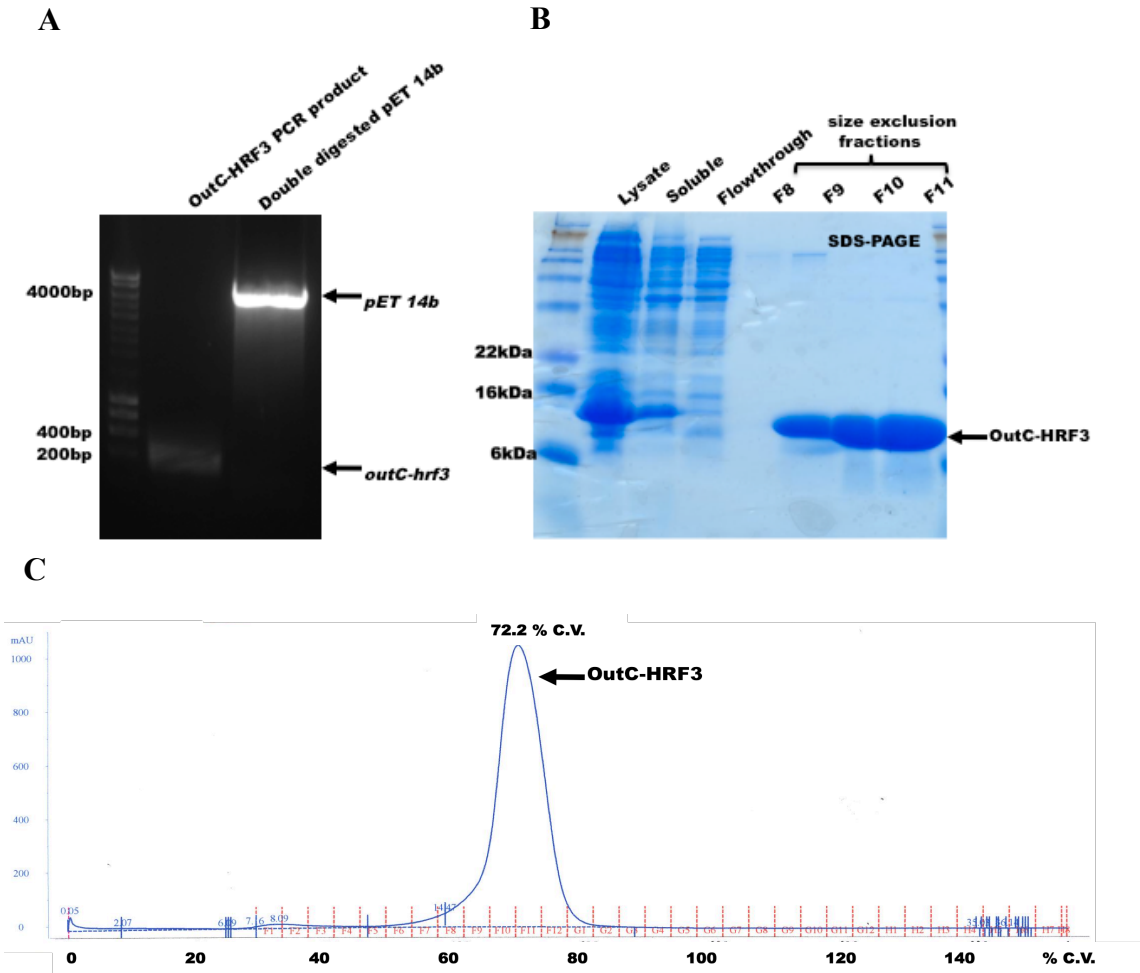


Figure 4.3 Cloning, expression and purification of OutC-HRF3. (A) Cloning of *outC-hrf3*. (B) Expression and purification of OutC-HRF3. (C) Gel-filtration trace of OutC-HRF3 on S200 HR 10/30 column.

4.3.2 OutC-HRF3 expression

The OutC-HRF3 expression was achieved in *E coli* BL21 (DE3). OutC-HRF3 encoding pET-14b vector was transformed into to BL21 (DE3) cells. Cultures (1 liter), inoculated by 5ml overnight culture, and were grown at 37^{o}C in Luria broth (LB) containing 500μg ampicillin until the absorbance at 600nm reached 0.6. The recombinant protein expression was induced by adding 1mM IPTG and growing for a further 4 hrs at 37^{o}C .

4.3.3 OutC-HRF3 purification

As the pET-14b expression vector produces recombinant protein with an N-terminal His-tag, a Ni-HiTrap affinity column (GE Healthcare) was used to purify the soluble protein. After loading the soluble protein fraction from the cell lysate, the column was washed with low pH low imidazole buffer and then thrombin-containing buffer.The OutC-HRF3 was produced with His tag cleaved off by incubate in thrombin protease containing buffer overnight. Protein was then eluted from the column by washing with buffer.

As shown in Figure 4.3B the protein eluted from the Ni column has relative high purity of OutC-HRF3 with the only visible contamination at around 70kDa. Since the size of the contamination is much greater than OutC-HRF3, size exclusion chromatography was used for further purification. After size exclusion chromatography the OutC-HRF3 fraction, as judged by SDS-PAGE, had a purity of at least 95%.

The OutC-HRF3 elutes from the S200 column at 72.2% column volumn (C.V.), consistent with a molecular weight of 9KDa. With the OutC-HRF3 calculated mass of 10.343kDa, the size exclusion column result suggests the OutC-HRF3 present in

solution as a monomer. On the other hand, OutC-HR with size of 18kDa elutes at 61.2% C.V. indicating the size around 35kDa. Since there is no evidence that OutC-HR exists in a dimeric form without the trans-membrane segment, the early elution of OutC-HR from the size exclusion column suggests the OutC-HR contains unstructured polypeptide that causes it to behave anomalously. In contrast to this, the OutC-HRF3, limited proteolysis fragment obtained from HR is more compact and structured.

To check that the recombinantly produced OutC-HR fragments maintain their structure as produced by proteolysis, SDS-PAGE and Native-PAGE were employed to compare both denatured and native states (Figure 4.4). The recombinantly produced OutC-HRF2 and OutC-HRF3 run similarly to the corresponding proteolytic fragments on both SDS-PAGE and Native-PAGE. The similarity on SDS-PAGE indicated they are similar in molecular weight. Whereas the similarity on Native-PAGE means they have similar charge to mass ratio and hints they fold in same way.

As observed in Figure 4.4, freshly produced OutC-HR showed a smeary band, indicating it is prone to degradation. The prepared OutC-HRF3 sample was found to be stable at 4^{o}C for at least 6 month.

4.4 Dynamic light scattering

Dynamic light scattering (DLS) is an established technique in the field of protein crystallography. It measures hydrodynamic size, polydispersity and aggregation of protein samples and gives a general idea of the ability of the protein to form crystals.

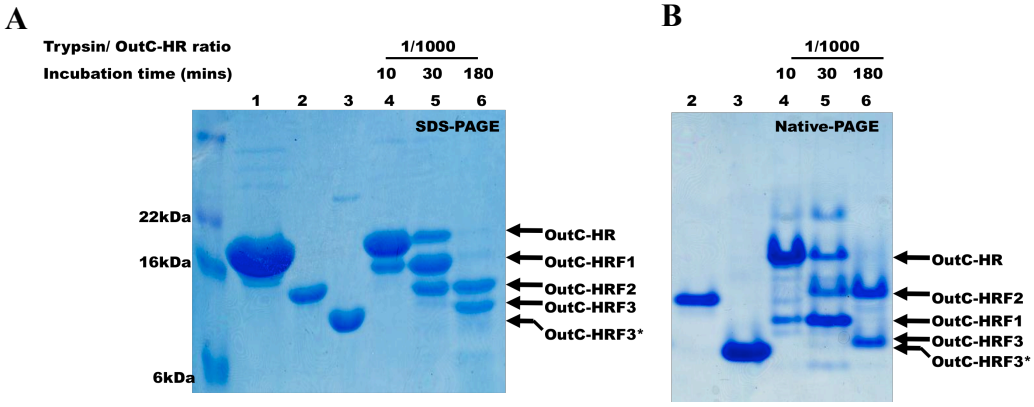
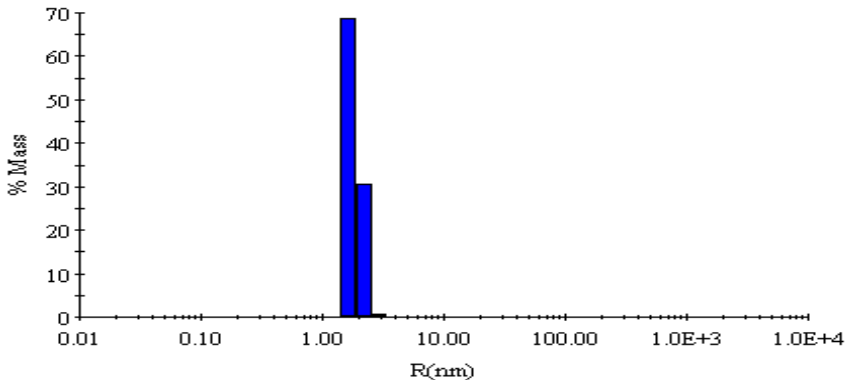


Figure 4.4 Comparision of recombinantly produced OutC-HRF2 and OutC-HRF3 and proteolysis fragments. (A) SDS-PAGE. (B) Native-PAGE. Lane 1, OutC-HR. Lane 2, purified pET14b OutC-HRF2 recombinantly produced protein. Lane3, purified pET14b OutC-HRF3 recombinantly produced protein. Lanes 4, 5 and 6 proteolysis of OutC-HR with trypsin. OutC-HRF3* is the recombinantly produced His tagged OutC-HRF3.

The OutC-HRF3 sample shows a polydispersity of 14.4%, which indicates OutC-HRF3 is monodisperse in solution (Figure 4.5). The average molecular mass calculated from the radius of the protein is approximately 14kD, which confirms that OutC-HRF3 is present in the solution as homogenous subunit. It is generally considered that protein with monodispersity of 20% or less is very likely to form crystals. The OutC-HRF3 DLS result suggests it can be crystallized.



	Radius(nm)	%Pd	MW-R(kDa)	%Int
Peak1	1.65211	14.4	14	85
Peak2	14.4235	20.9	1734	15

Figure 4.5 OutC-HRF3 DLS summary. 14.4% polydispersity, radius 18.5 nm and molecular weight 14kDa.

4.5 OutC-HRF3 solution structure

OutC-HRF3 protein was put into crystallization trails. Crystal screens was carried out with commercial screens: including Hampton Research screen I and II, Wizard crystallization screen I and II, using mannual hanging drop as well as robotic sitting drop technique. Crystallization temperature was chose at 18^{o}C, which is consistent with DLS experiment condition. However, no crysal was obtained. The solution structure was therefore solved by NMR spectroscopy.

4.5.1 Determining the architecture of OutC-HRF3 using NMR spectroscopy

To ensure that the OutC-HRF3 maintained the structure as in its functional form, the OutC-HRF3 ^{1}H-^{15}N HSQC spectra was collected and compared with the OutC-Peri spectra (Firgure 4.6). The ^{1}H-^{15}N HSQC of OutC-HRF3 reveals a set of well dispersed peaks, where almost all peaks overlay well with OutC-Peri peaks. Protein chemical shifts are very sensitive to local electron environment and the excellent overlap between OutC-HRF3 and OutC-Peri shows that the HR domain structure is independent of the rest of the periplasmic region of OutC. There will be very few or no contacts between the HR and PDZ domains of OutC. This result is consistent with the "beads on a string" model of the domain architecture of the periplasmic region of OutC proposed by Login et al. (Login, Fries et al. 2010). Most importantly, this result confirms that OutC-HRF3 retains the functional structure of OutC-Peri.

The OutC-HRF3 ^{1}H-^{15}N HSQC spectra was assigned. The construct contained an additional 7 residues in a tag (see Appendix 1 for sequence). The 7 residues are present at the N-terminus of the protein and are consequence of the thrombin cleavage site. The OutC-HRF3 backbone amide proton was fully assigned, except the 7 residues in the tag and 5 prolines in the sequence (Figure 4.7).

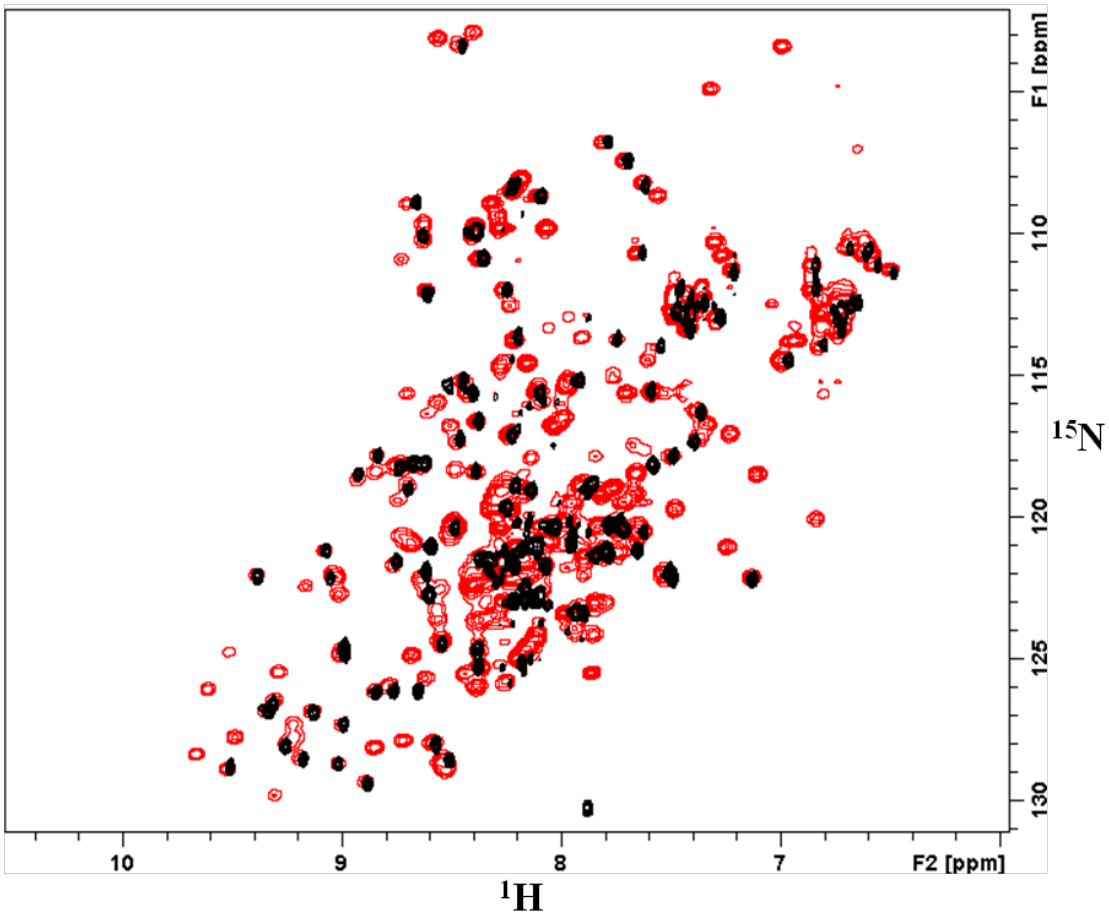


Figure 4.6 Comparison of OutC-HRF3 and OutC-Peri spectra. 2D 1H-15N HSQC of OutC-HRF3 (black) and 2D ^{1}H-^{15}N HSQC of OutC-Peri (black). Both spectra acquired at 15°C on Bruker 600MHz in 20mM Tris (pH7.0) and 10% $^{2}H_2O$.

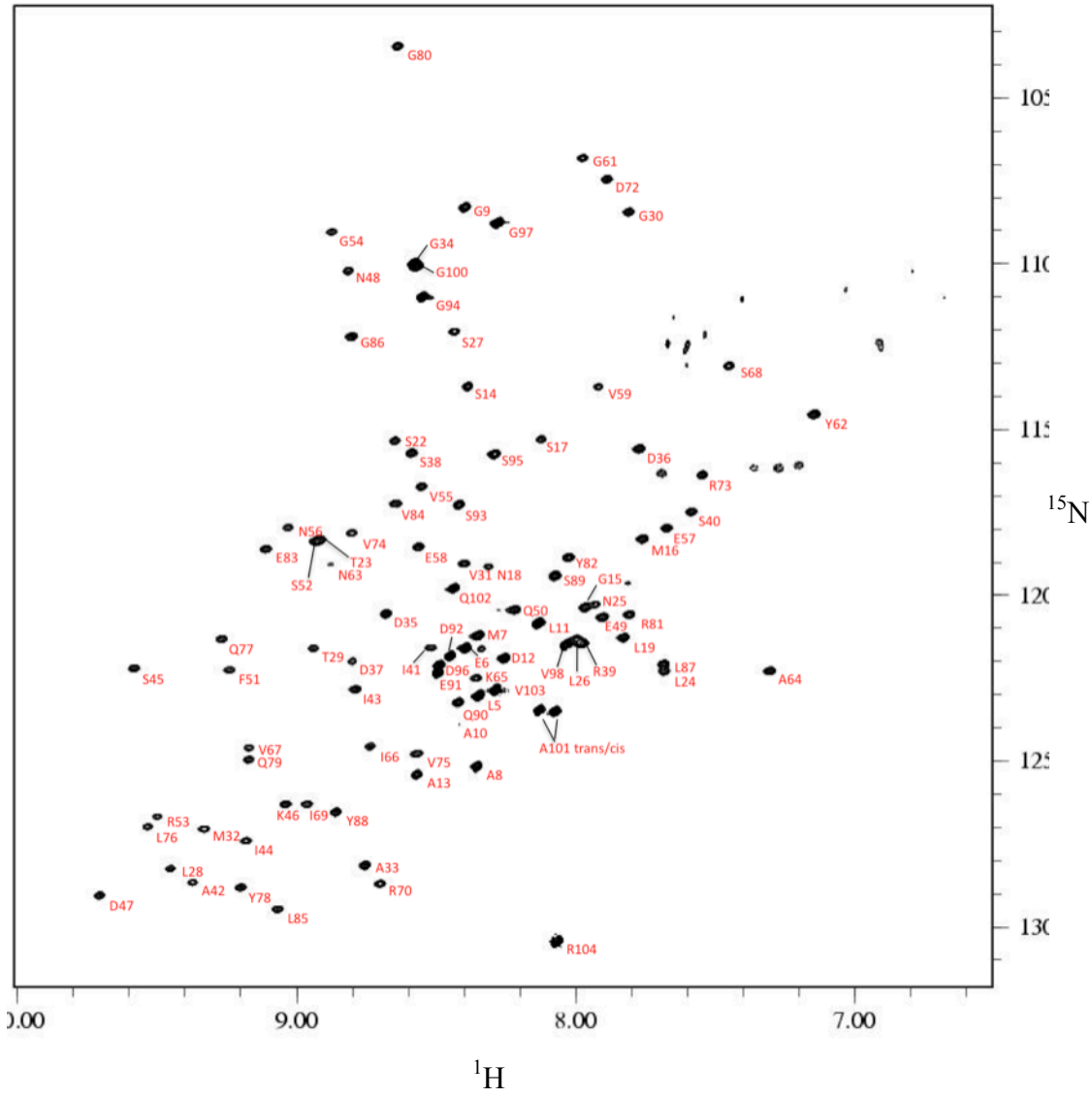


Figure 4. 7 ^{1}H-^{15}N HSQC spectrum of OutC-HRF3. The spectrum was acquired on a Bruker 600MHz NMR spectrometer at 15^{o}C in 20mM Tris (pH 7.0) and 10% $^{2}H_2O$.

4.5.3 OutC-HRF3 secondary structure prediction

The OutC-HRF3 protein structure determination began with a secondary structural prediction using TALOS+ (Shen and Bax 2007; Shen Y 2009). Protein backbone chemical shift is exquisitely sensitive to local conformation. By using the inverse of this relation, TALOS+ compares the input protein sequence with a data base of 200 proteins and selects the 10 best matches to the secondary chemical shifts of given residual in a target protein along with one extra residue on each side (a residual triplet). If there is a consensus ϕ and ψ angle among the 10 best database matches, then database triplet structures is used to predict ϕ and ψ angles of the central residue.

TALOS+ incorporates the assignments of ^{15}N, ^{1}H, $^{13}C\alpha$, $^{13}C\beta$, $H\alpha$, $H\beta$ and ^{13}C chemical shifts of each amino acid residue for secondary structure prediction. It also uses the Random Coil Index (RCI) derived method to predict S^2 backbone chemical shifts. OutC-HRF3 secondary structural prediction reveals a very short helical structure at the N-terminus and followed by six β-strands. The RCI S^2 reflects the internal picosecond-nanosecond motion (Berjanskii and Wishart 2005). Since S^2 value significantly varies for the mobile region, the predicted OutC-HRF3 S^2 value suggests the N-terminal 12 residues and C-terminal 18 residues are flexible.

OutC-HRF3 shows a structure formed of predominantly β-sheet structure: with short helix at N-terminus and followed by 6 beta-strands. This is consistent with the secondary structure prediction done by J-pred (http://www.compbio.dundee.ac.uk/www-jpred/), which successfully predicted the 6 beta-strands. However, the helical region is not highly conserved across species.

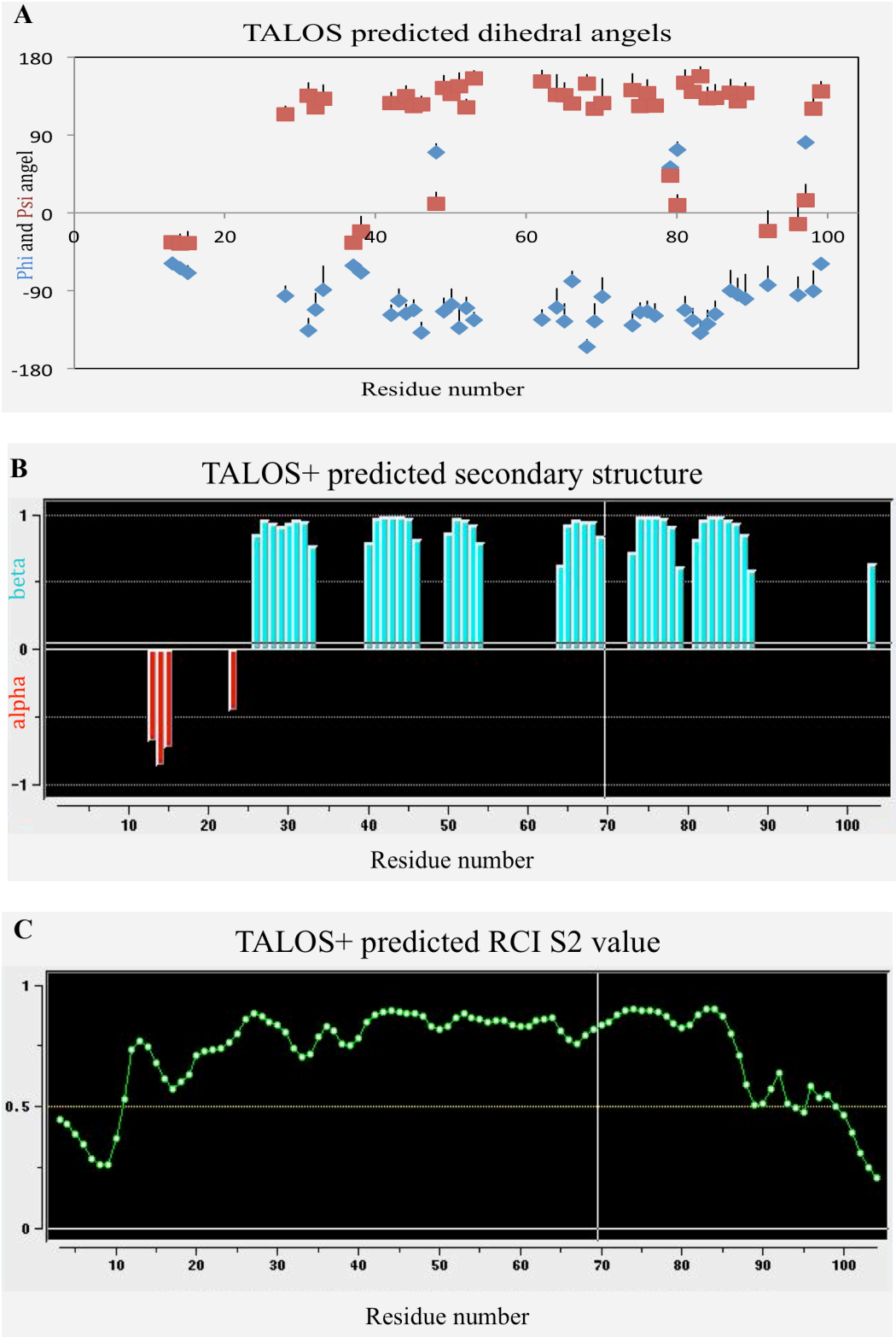


Figure 4.8 Secondary structure prediction of OutC-HRF3. (A) TALOS+ predicted Phi Psi angels for residues which significant matches of Phi and Psi angels were made. (B) Secondary structure predicted by Phi and Psi torsion angles. (C) TALOS+ predicted RCI-S^2 (upper panel) and secondary structure (lower panel) for OutC-HRF3 (cyan beta-sheet; red, helix).

For the structural calculations presented in the following sections, TALOS derived ϕ and ψ dihedral angle restraints were incorporated with a TALOS factor of 2 and TALOS error of 20 in the final ARIA structure calculation.

4.5.5 Relaxation studies

The TALOS+ predicted S2 value gives an idea of the internal dynamics of OutC-HRF3, more detailed relaxation studies were carried out to further investigate its hydrodynamic properties.

The T_1, T_2, and ^{1}H-^{15}N-NOE relaxation profiles of OutC-HRF3 are shown in Figure 4.9. In each relaxation profile, missing residues correspond to prolines, overlapped peaks or peaks with extremely weak intensity. The structured domain of OutC-HRF3 is defined in the T_2 and ^{1}H-^{15}N-NOE relaxation profiles, as being formed by residues 22-88. This agrees with TALOS+ predicted secondary structure region and S2 value. In this structured region of OutC-HRF3 the T_1 relaxation time is ~784.24±2.34 ms, T_2 relaxation time is ~87.05±0.436 ms, ^{1}H-^{15}N-NOE relaxation is 0.72. Using these relaxation data, the correlation time of OutC-HRF3 measured at 15 ^{o}C is 8.6 ns. The correlation time of 8.6ns indicates that OutC-HRF3 behaves in solution as an isolated subunit, agreeing with size-exclusion and DLS data.

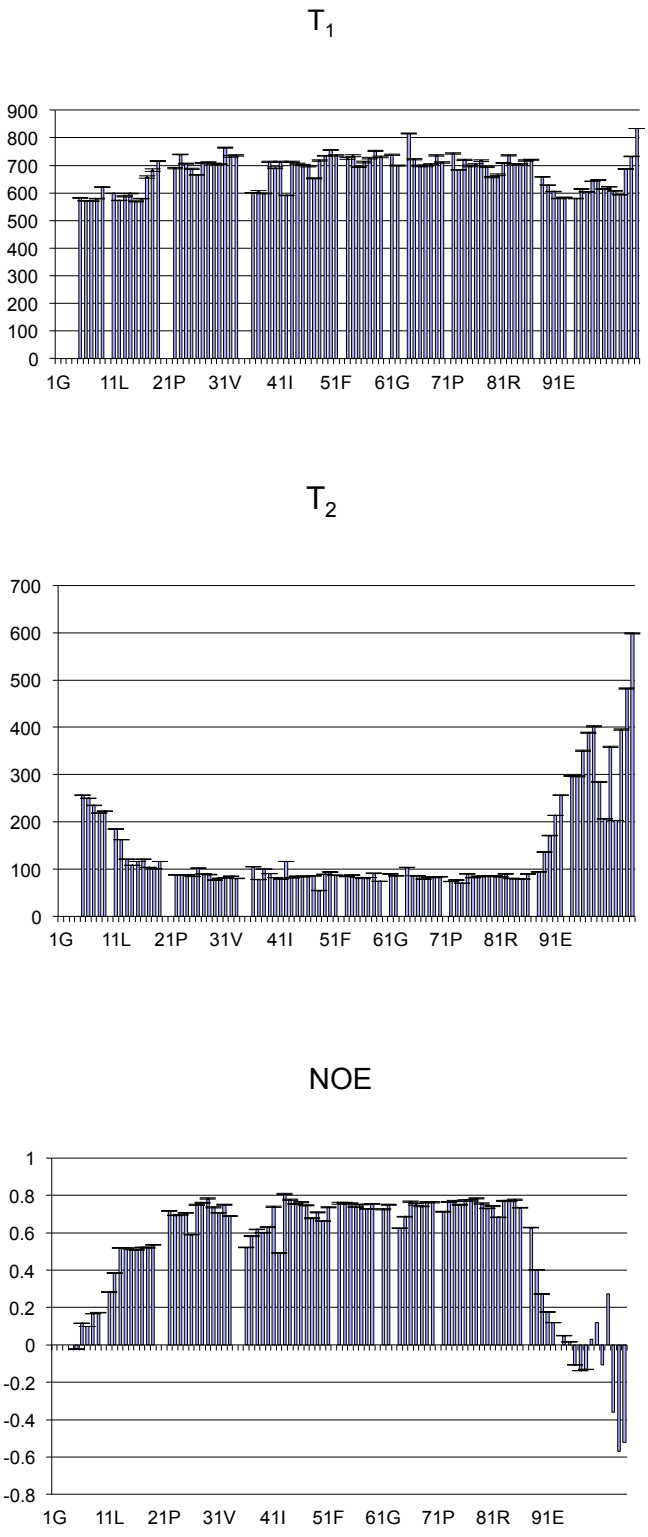


Figure 4.9 T_1,T_2 and NOE relaxation value of OutC-HRF3. Acquired at Varian 600MHz at 15^{o}C in 20mM Tris (pH 7.0) and 10% $^{2}H_2O$.

4.5.6 CS-ROSETTA structure prediction

It has long been recognized that chemical shifts are strongly influenced by local conformation. To date, interpretation of isotropic chemical shifts in protein structure is largely based on empirical correlations gained from the mining of protein chemical shifts deposited in Biological Magnetic Resonance Data Bank (BMRB), in conjunction with the known corresponding 3D structures. Chemical-Shift-ROSETTA (CS-ROSETTA) is a robust protocol to exploit this relation for the *de novo* protein structure generation for proteins no larger than 15kDa, by using as input parameters the 13CA, 13CB, 13C', 15N, 1HA and 1HN NMR chemical shifts. These shifts are generally available at the early stage of conventional NMR structure determination procedure. However, CS-ROSETTA has a tendency to slightly lengthen the elements of secondary structure and include residues that were clearly disordered as judged by the NMR data. This is due to the program energy selection process favors the formation of intra-molecular hydrogen bonds, and a tendency to generate secondary structure for disordered region is therefore not surprising.

The OutC-HRF3 structure was predicted using CS-ROSETTA with the chemical shifts available at the early stage of the structure determination. 10,000 structures were generated and the best 1000 structures selected by energy. The energy of the best 1000 structure is plotted against the lowest energy structure (Figure 4.10). The root-mean-square deviation (RMSD) of the 10 lowest energy structures is within 3Å, which suggest the structures predicted by CS-ROSETTA converged successfully and are reliable. The best 200 structures are very similar and share same structure fold.

CR-ROSETTA predicts OutC-HRF3 adopts a beta-sandwich fold (Golovanov et al

2006): with 2 beta sheets, each composed of three hydrogen-bonding anti-parallel strands (Figure 4.11). When using this predicted OutC-HRF3 structure as input to Dali database server (Holm and Rosenstrom 2010), the only hit with similar structure was PilP, with Z score of 6.4 and backbone RMSD of 1.9 Å. PilP is a pilotin from typeIV secretion system. Although the exact function of PilP is not understood, it occupies a similar location to OutC-HRF3 located on the inner-membrane and reaching into periplasma to interact with the secretin in the outer-membrane (Drake, Sandstedt et al. 1997). As OutC-HRF3 shares a similar function and interacts with T2SS secretin, the predicted OutC-HRF3 structure is highly likely to be correct and could be the first PilP-like structure discovered (Figure 4.12).

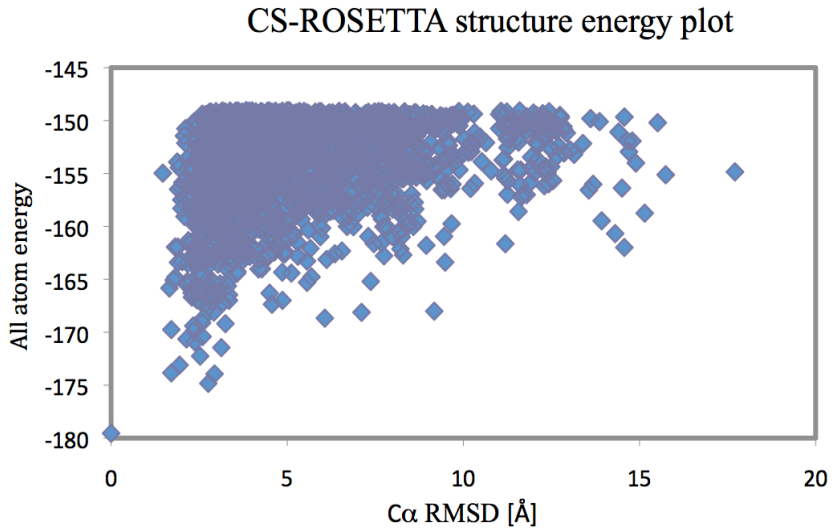


Figure 4.10 CS-ROSETTA structures energy plot. The 2000 lowest energy structures were plotted according to the lowest energy structure. 10 lowest energy models all differ by less than 3Å.

The CS-ROSETTA predicted OutC-HRF3 structure is show in Figure 4.11. The protein adopts a beta-sandwich-type fold. A short 3_{10} helix is formed at the N-terminus, which is also found in PilP (Golovanov et al).

In the OutC-HRF3 predicted structure the characteristic **β**4 strand found in PilP, which is also revealed as highly conserved strand in the type IV PilP equivalent proteins in

other species, is replaced by a more flexible loop. The two key residues involved in forming hydrophobic core are I113 and V144. I 133 is conserved between the OutC-HRF3 and PilP . V144 substituted by a similar hydrophobic residue Ile. The CS-ROSETTA predicted structure was later confirmed to be correct using the standard NOE based structure determination procedure (detailed below).

4.5.7 Hydrogen bond restraints

Hydrogen bonds stabilize protein structure. It is very informative for structure calculation and refinement. Indirect hydrogen bond measurement by quadrupole coupling constant (Liwang and Bax 1997) was unsuccessful due to rapid backbone amide exchange with solvent deuterons for almost all residues. Fortunately, OutC-HRF3 hydrogen bonds can be measured by direct trans-hydrogen bond (N-H$\cdots$O=C) scalar coupling (Cordier and Grzesiek 1999). Compared to the indirect D_2O exchange method, this method is relatively difficult to perform and gives few hydrogen bonds but the result is more precise.

In Figure 4.13 the OutC-HRF3 standard HNCO spectra is overlaid with the results of the OutC-HRF3 long-range $J_{NC'}$ HNCO spectra. The chemical shifts of $NH_{(n)}$ and $CO_{(n-1)}$ on the standard HNCO spectra allow identification of the residue, n. If the residue HN is involved in hydrogen-bonding a CO group, the hydrogen-bonding partner CO can be identified from the chemical shifs of the NH and CO on the long range $J_{NC'}$ HNCO experiment. The NH shift will be in common and the different CO shift for the hydrogen binding carbonyl allows identification of the long-range CO. 28 hydrogen bonds were measured in OutC-HRF3 and these are listed in Appendix 3. Hydrogen bond restraints were included in the final structure calculation.

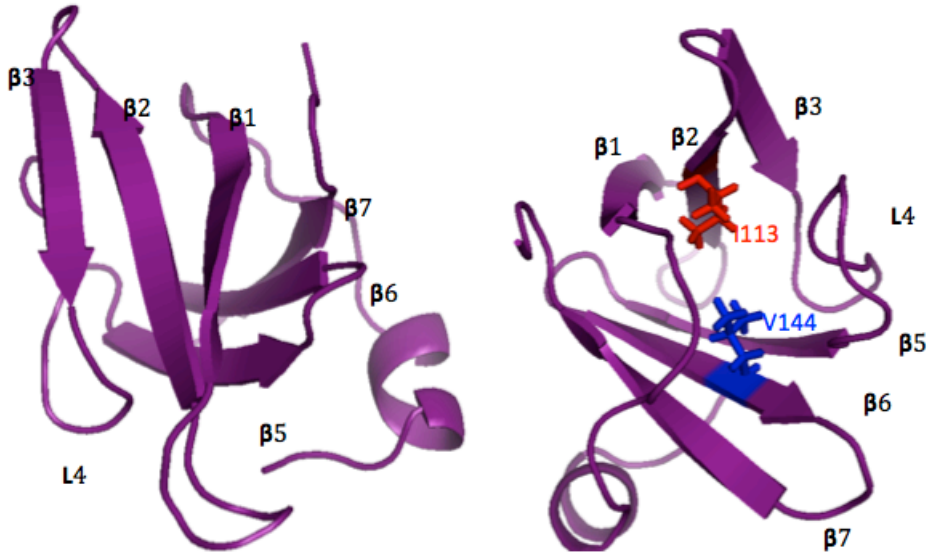


Figure 4.11 Two views of CS-ROSETTA predicted OutC-HRF3 structure. The hydrophobic core residue I 113 and V144 are colored in red and blue. The structure comprises a beta-sandwich fold.

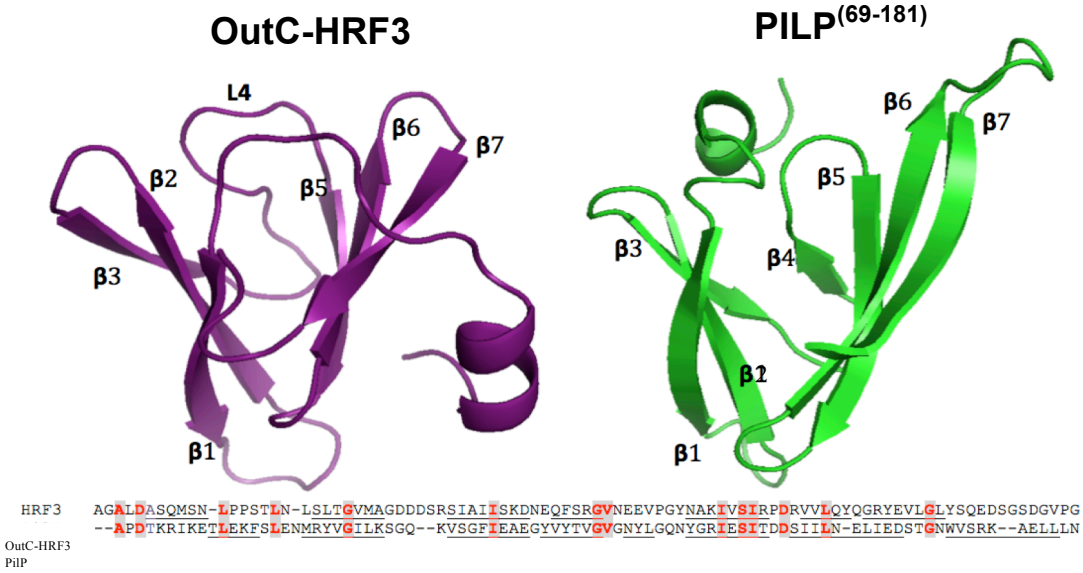


Figure 4.12 Comparison of OutC-HRF3 and PilP structure. OutC-HRF3 shows a longer linker between the 3_{10} helix and β1 strand, shorter β6 and β7 strand and b4 is replaced by a loop (L4).

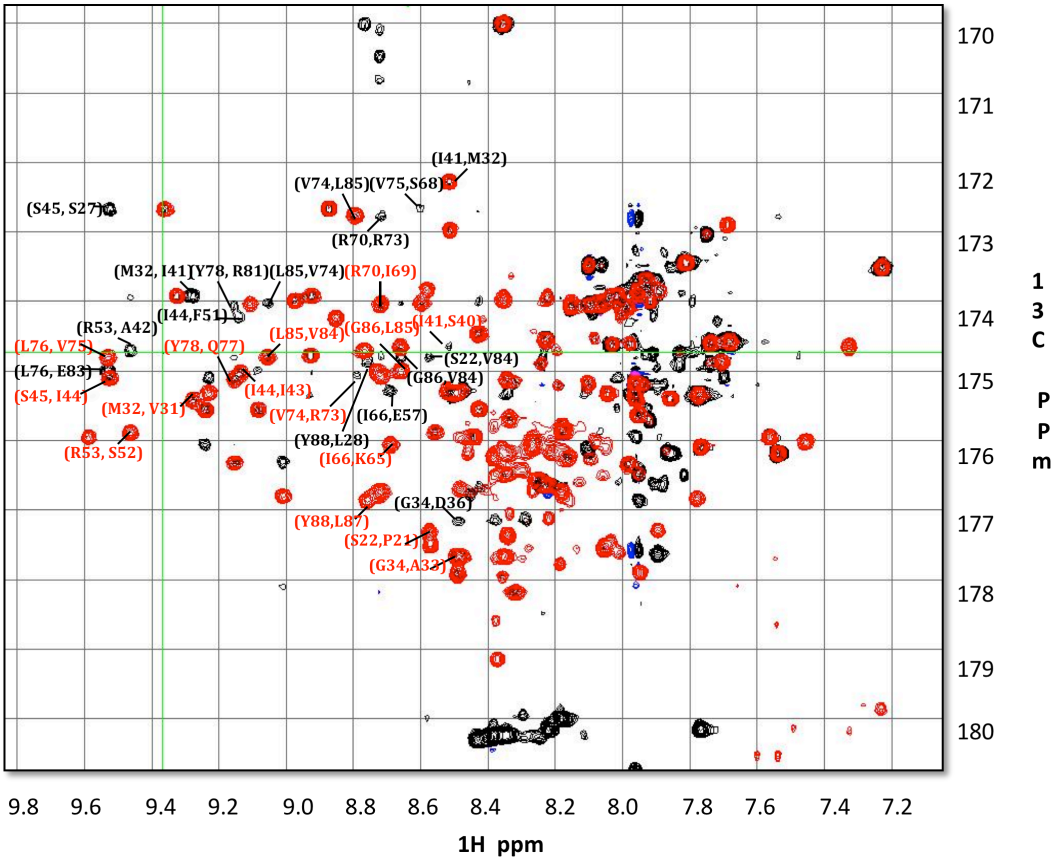


Figure 4.13 OutC-HRF3 hydrogen bond measurement. OutC-HRF3 standard HNCO experiment (in red), showing omega bond. OutC-HRF3 long-range $J_{NC'}$ HNCO experiment (in black), showing hydrogen bond between residues. Identified peaks are labeled as (HN chemical shift, CO chemical shift).

4.5.8 Residual dipolar coupling

Residual dipolar coupling (RDC) measured in diluted liquid crystalline phase provides important restraints for molecular structure determinations by NMR spectroscopy, which cannot be obtained otherwise. Dipolar coupling is a through-space interaction, which arises between any two magnetically active nuclei. Residual dipolar coupling is a useful method to provide independent information on how each dipole is orientated related to the molecular coordinate frame and in turn, to each other. It depends on distance, orientation and dynamics. As a result of Brownian motion, dipolar couplings average to zero under isotropic conditions and are only observed under anisotropic conditions.

For two dipole-coupled nuclei, A and B, the observable dipolar coupling in solution, D_{AB} can be expressed as:

$$D_{AB}(\theta,\phi) = A_a^{AB}\{(3\cos^2\theta - 1) + 3/2R(\sin^2\theta \cos 2\phi)\}$$

A_a^{AB} and R are the axial and rhombic components, respectively, of the molecular alignment tensor, A in the principle coordinate frame.

To use RDC in any type of structure refinement, good estimates for A_a^{AB} and R in must be available. The A_a^{AB} and R value for OutC-HRF3 were calculated to be 0.5 and 7 using maximum likelihood method (Warrena and Moore 2002).

As shown in Figure 4.14 the measured HN-N residual dipolar coupling in OutC-HRF3 can be plotted as a linear function. This indicates the calculated structure agree with the RDC measurements. The RDC measured with greater standard deviation is normally located on the flexible regions of the structure, such as loops connecting the strands.

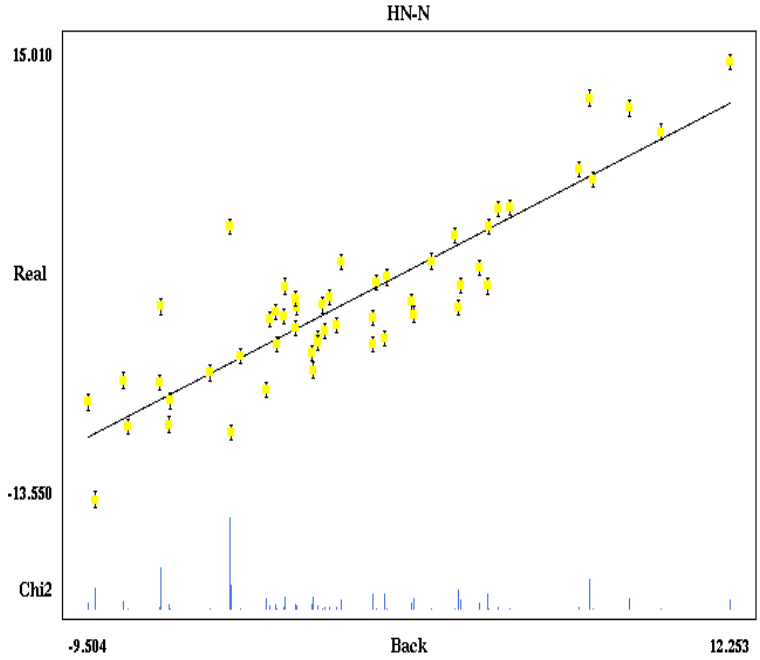


Figure 4.14 Correlations between calculated OutC-HRF3 NMR structure and experimental measured RDC.

4.5.9 Resonance assignment of OutC-HRF3

Though analysis of the various triple resonance spectra outlined in Section 3.1 the assignment of OutC-HRF3 backbone and side chain resonances was achieved. In total, OutC-HRF3 contains 1373 potentially assignable atoms. Assignment of several subsets of atoms such as the side chain NH_2 groups of Arg residues, Lys NH_3^+ groups and in many cases the hydroxyl protons of Ser and Thr residues, were not possible.

The overall OutC-HRF3 assignment statistics are shown in Table 4.1. The core region revealed by the relaxation study and including the β-sandwich fold, residues 22-87, has the highest proportion of assigned atoms (83.7%).

OutC-HRF3	Total number of atoms	Number of unassigned atoms	% Assignment
Full length (1-104)	1373	304	77.8
Core residues(22-87)	918	150	83.7

Table 4.1 OutC-HRF3 assignment statistics.

4.5.10 NMR structure calculation

ARIA1.2 was used for OutC-HRF3 structure calculation. TALOS calculated dihedral angles (Appendix 5), H-Bond (Appendix 3), RDC (Appendix 4), and manually identified unambiguous NOE files together with ^{1}H-^{15}N-NOESY and ^{1}H-^{13}C-NOESY peak lists were used for structure calculation.

The NOE intensity is inversely proportional to the 10^6 power of the distance, for the residues in flexible region, the NOEs measured favor the conformation that gives strong NOEs and could not represent the dynamics of the flexible region. Therefore, NOEs involved in the flexible region residues were excluded from the calculation by deleting

the flexible region assignment. According to relaxation data, residues 22 -87 behave as the stable structural core of the protein. The NOE peak intensities from 13C-NOE and 15N-NOE spectra were used in the ARIA calculation with spectral calibration (qcali=ture in run.cns).

4.5.11 OutC-HRF3 solution structure evaluation

OutC-HRF3 NMR relaxation measurement revealed that the structured region comprised 65 residues, OutC-HRF3$^{22-87}$. A high percentage sequence-specific NMR signal assignment was achieved using a combination of 3D triple resonance experiments using ^{15}N, ^{13}C-labeled OutC-HR, and 3D NOSEY-HSQC, TOCSY-HSQC using ^{15}N-labeled protein. The 1912 experimental nuclear Overhauser effect (NOE) distance restraints were obtained using ^{13}C or ^{15}N -edited NOE spectroscopy (NOESY) spectra, and 76 φ and ψ torsion angle restraints were derived from secondary chemical shift analysis using the TALOS algorithm. Twenty eight hydrogen bonds were experimentally measured. The structure of OutC-HRF3$^{(22-87)}$ was determined with high precision: the pairwise RMSDs for the 20 conformers of the final ensemble were 0.46Å and 1.26 Å for the backbone and heavy atoms respectively. The structural quality, as assessed by the standard for the backbone. NMR statistics summarised in Table 4.2. It is significant that even the loop regions connecting the anti-parallel β-strands are well-determined in this structure (Figure 4.16). The only disallowed residue from PROCHECKnmr is Asp36, located on a sharp turn between β1 and β2. In the core structure region, the calculated final 20 structures and best rosetta model has a RMSD within 1.5 Å.

The architecture of the central 65 residues is β-sandwich-like comprising two three-

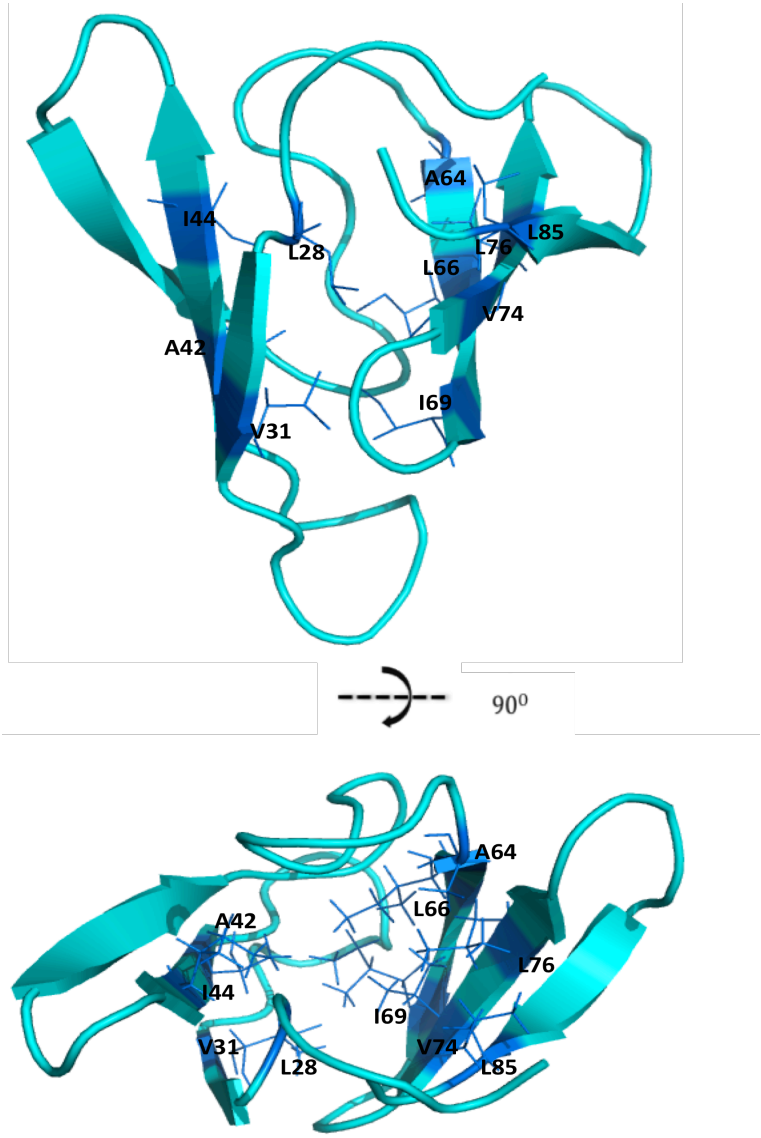
stranded (up-down-up) anti-parallel β-sheets with short hairpin loops connecting adjacent β-strands within the two sheets and a longer loop between the two sheets (Figure 4.15). The angle between the strands in the two β-sheets is approximately 45^0.To both the N- and C- terminal ends of the β-sandwich there are approximately 20 residues which are less regular in structure although to the N-terminal end of the structure there is a short 3_{10}-helix.

The two β-sheets composing the HR domain seem largely independent and are connected through several hydrophobic interactions (Figure 4.15), involving Val31, Ala42, Ile44, Ile66, Ile69, Val74 and Val76.

In the OutC-HRF3 folded state there is quite an uneven charge distribution. There is a large hydrophobic patch on one side of the protein, mainly formed by β1 and β2 strands (Figure 4.17). This provides a large platform for hydrophobic interaction. The other side of the OutC-HRF3 displays a mix of positive, negative and hydrophobic patches. The function of these feature during the assembly and action of T2SS has yet to be discovered.

	OutC-HRF3
NMR distance and dihedral constraints	
Distance constraints	
Total NOE	1912
Intra-residue	629
Inter-residue	
Sequential (|i-j|=1)	373
Medium-range (|i-j| ≥ 2 ≤ 4)	166
Long-range (|i-j| ≥ 5)	744
Hydrogen bonds (experimental measured)	28
Hydrogen bonds(observed in over 50% of structures)	64
Total dihedral angle restraints (TALOS)	76
Φ	38
Ψ	38
Structure statistics	
Violations (mean ± SD)	
Distance constraints (>0.5) (Å)	1.6±0.55
Dihedral angle constraints (>5)(deg)	1.9±1.9
Deviations from idealized geometry	
Bond lengths (Å)	0.005± 0.0001
Bond angles (deg)	0.671± 0.011
Impropers (deg)	0.748± 0.024
Average pairwise r.m.s.d. (Å)[a]	
Backbone	0.46±0.11
Heavy	1.26±0.20
RMSD from the mean structure (Å)[a]	0.33±0.07
Residues in most favored regions [A,B,L][b]	60.70%
Residues in additional allowed regions[a,b,l,p][b]	33.90%
Residues in generously allowed regions [~a,~b,~l,~p][b]	3.60%
Residues in generously disallowed regions [~a,~b,~l,~p][b]	1.80%

Table 4.2 Experimental data and structural statistics for 20 energy minimised structure of OutC-HRF3 a. Pairwise RMSD and RMSD from the mean structure were calculated among the 20 lowest energy structures from 50 refined structures (amino acid residues 22 to 87).b Structure quality was analyzed with PROCHECK-NMR over structured regions (amino acid residues 22-87).




```
                  10        20        30        40        50        60        70        80
E.chrysanthemi    AGALDASQMSNLPPSTLNLSLTGVMAGDDDSRSIAIISKDNEQFSRGVNEEVPGYNAKIVSIRPDRVVLQYQGRYEVLGL
E.carotovora
SAVNDAALSGDIPLSSLNISLTGVLASEDAKRSIAIIAKDSQQYSRNVGDAVPGYEAKIVTISADRVVLQYQGRYEALHLK.pneumoniae
PAPTDATHLNQVPVSSLKLRLTGLLASSNPARSIAIIEKGNQQVSLSSGDTIPGYDARIVAILPDRIIVNYQGRKEAILL
V.cholerae        APVVEQPVVVDAPKTRLSLVLSGVVASNDAQKSLAVIANRGVQATYGINEVIEGTQAKLKAVMPDRVIISNSGRDETLML
E.coli            RAQETMPSLSNDLLSGEDLDVRGILYSSVAEHSVAIFAHNNRQFSLSVGEKVPSYDATISAIFSDHIVINYQGKTVSLPL
                                    β1        β2        β3        β4        β5        β6
```

Figure 4.15 Structure of OutC-HRF3$^{22-87}$. (A) Structure of OutC-HRF3 in ribbon representation with hydrophobic core residues shown in stick representation. (B) Family sequence alignment of GspC-HR Position of the secondary structure as indicated, with conserved residues in red background, semi-conserved residues in red and hydrophobic core residues underlined.

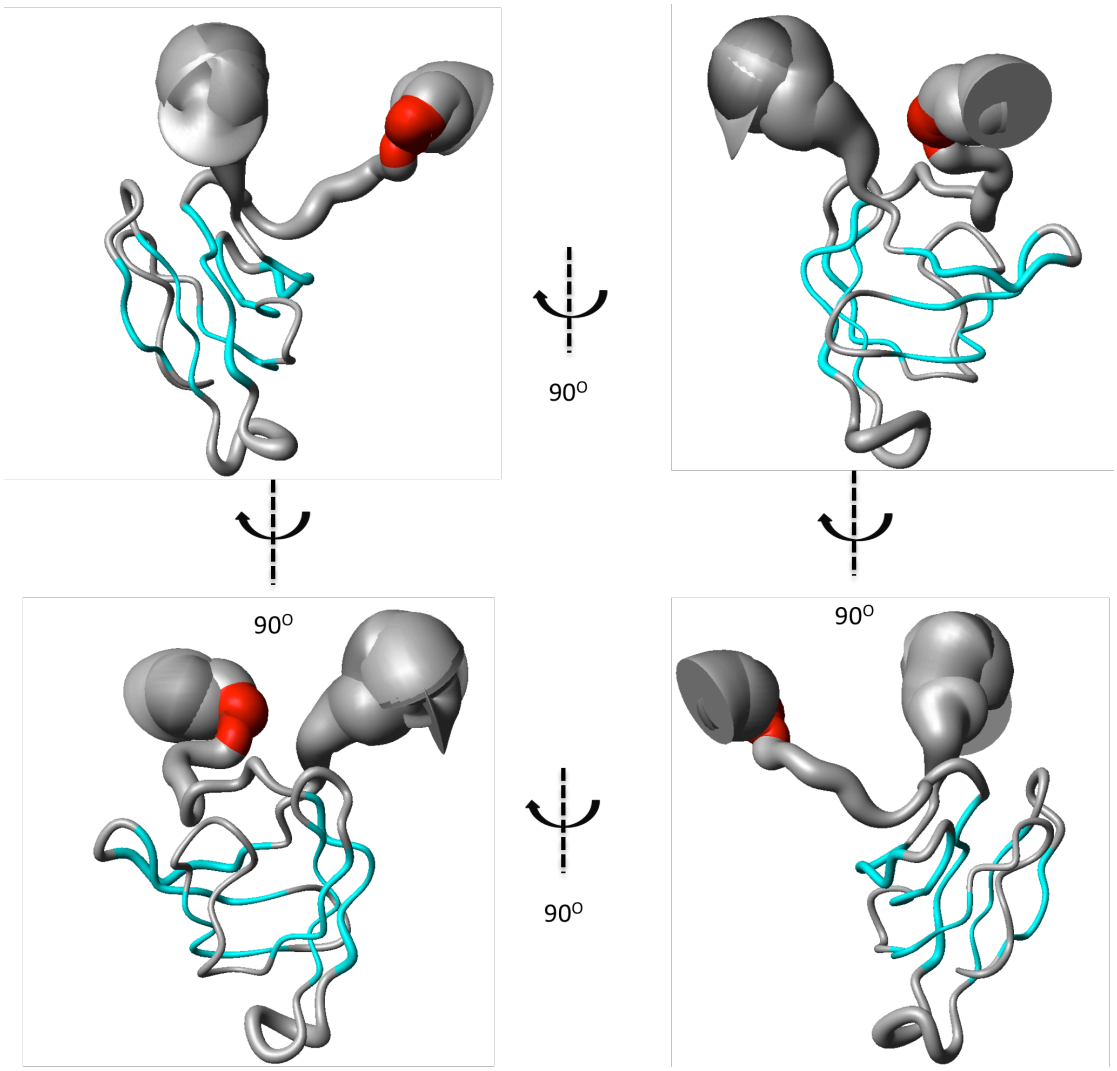


Figure 4.16 Sausage representation of best 20 calculated OutC-HRF3 structures. Secondary structure is shown in color, α-helix in red, β-sheet in cyan. The diameter of the sausage reflects the dynamics of the protein in solution, which is plotted according to the residue Cα T2 relaxation profile.

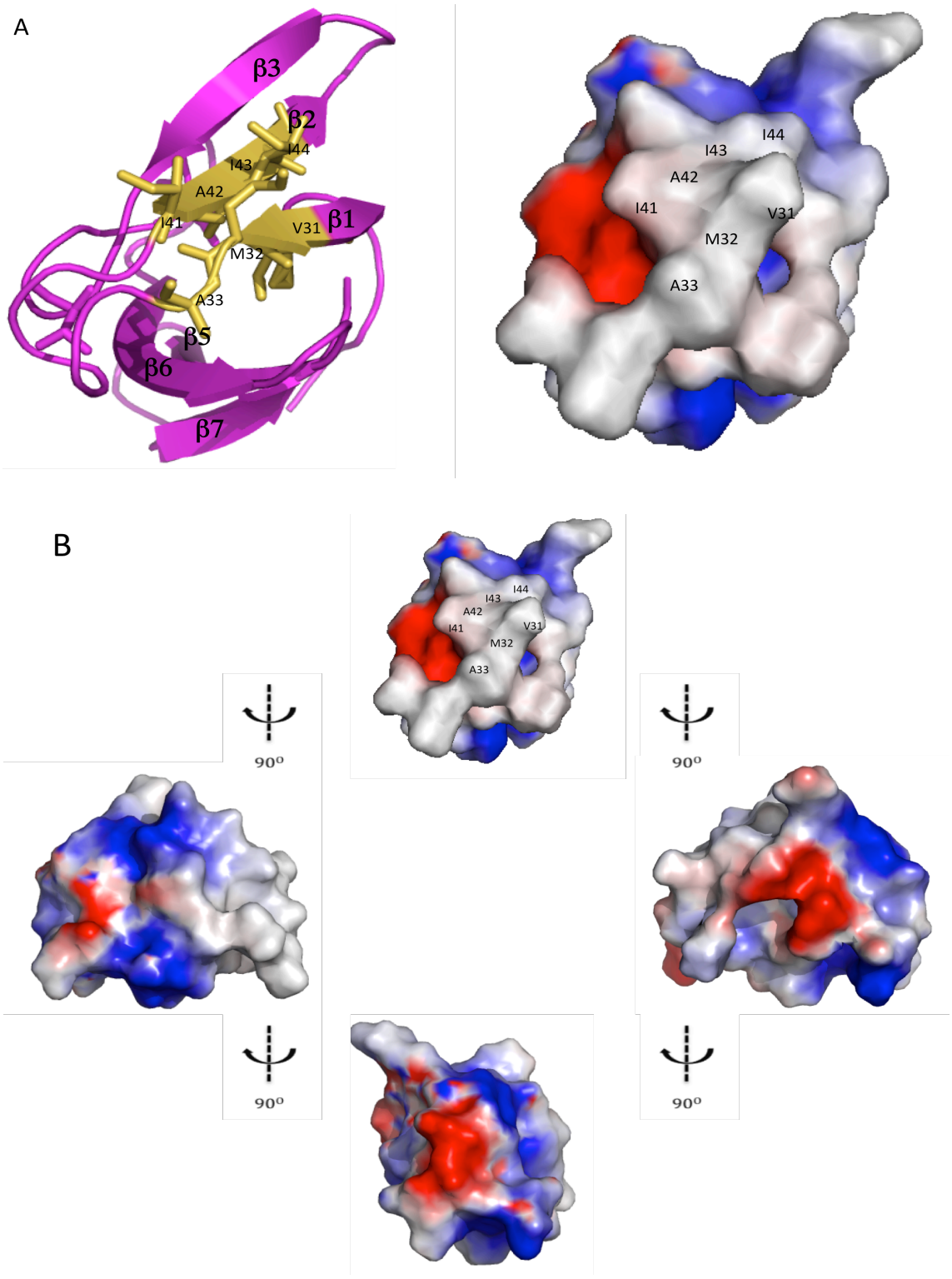


Figure 4.17 OutC-HRF3 electrostatic distribution. (A) β1 and β2 strand of OutC-HRF3 form a continuous hydrophobic surface. (B) OutC-HRF3 electrostatic distribution (shown in arbitrary PyMOL unit).

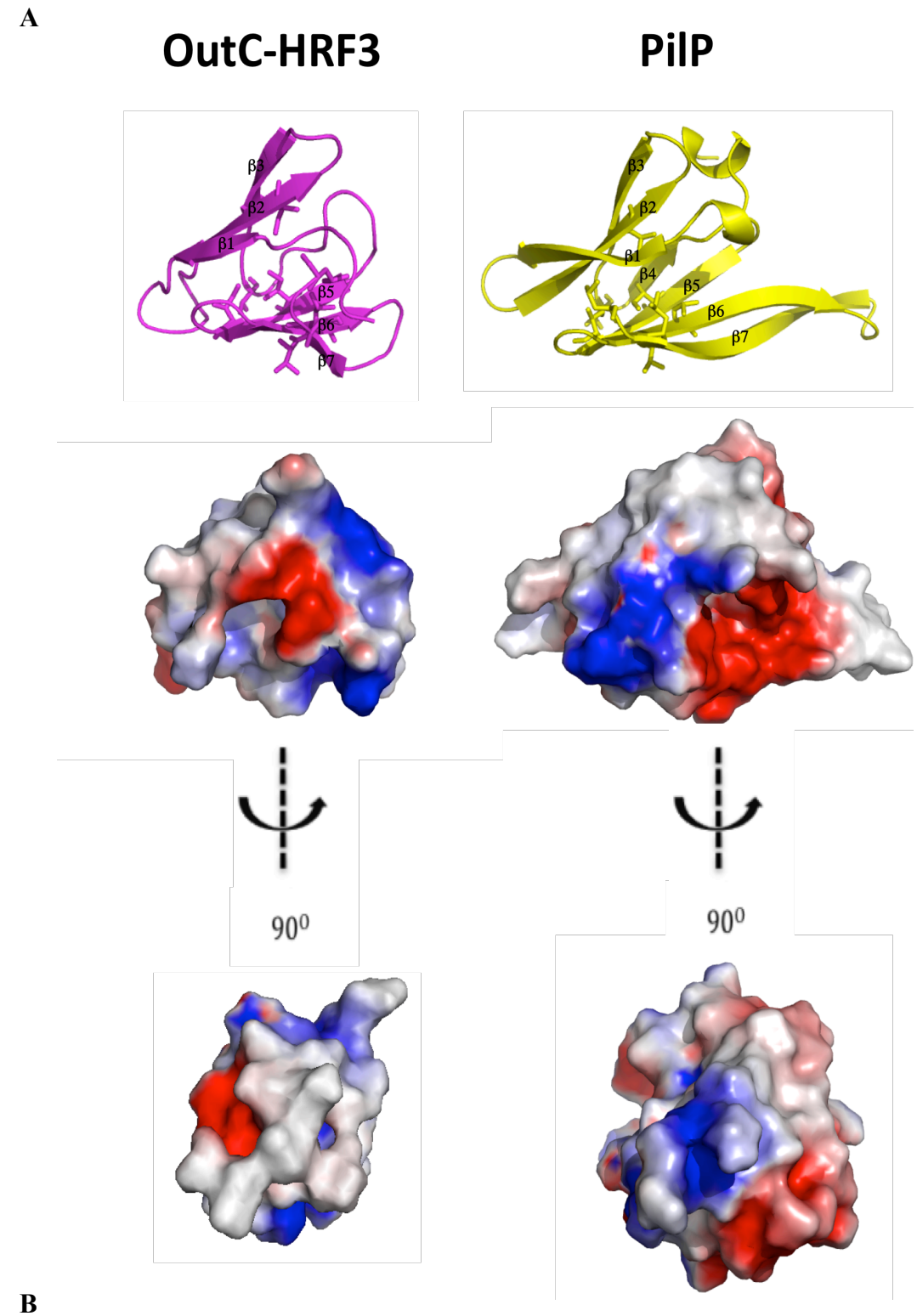


Figure 4.18 OutC-HRF3 and PilP comparisons. (A) Secondary structure, surface, surface electrostatic distribution comparisons , shown in arbitrary PyMOL unit. The conserved residue is shown in stick representation. (B) Sequence alignment, with secondary structure highlighted in colour.

4.6 Summary and discussion

In this chapter, the OutC-HR solution structure was solved by NMR spectroscopy. A β-sandwich fold was revealed. GspC-HR is the second structure having this fold. The other protein sharing this fold is the inner-membrane lipoprotein PilP from the type IV secretion system. PilP shares many similarities with GspC-HR including its location at the inner membrane, extending into the periplasm, interacting with an outer membrane secretin. They share a hydrophobic patch between strands β1 and β2 . It is plausible that PilP and GspC interact with their secretin's in a similar way. At the moment, there is no molecular details of how PilP interacts with its cognate secretin. However, both the GspC-HR and PilP interaction with secretin are essential for assembly of a functional secretion system. The interaction between OutC-HRF3 and its secretin OutD is further investigated in the following chapters.

Chapter 5 Study of the interaction between OutC-HRF3 and OutD-N0

5.1 Overview

The OutC homology region (HR) is the only highly conserved region of GspC across all species. It plays a key role in maintaining the functional T2SS. The OutC-HR has been shown to interact with two sites on the OutD N-terminal region, docking the inner membrane components with the outer membrane secretin subunits (Login, Fries et al. 2010). It may also interact with secreted proteins. When the PDZ domain of OutC is deleted, the system loses the ability to secrete pectate lyases (Bouley, Condemine et al. 2001) but the system can still secrete other proteins revealing that the PDZ domain is not the only domain involved in the recognition of proteins to be secreted. When the OutC-HR is deleted the system loses the ability to secrete all proteins. This may due to the involvement of OutC-HR in the assembly of the inner and outer membrane components. In this Chapter the interaction between OutC-HR and OutD N-terminal domains is studied. OutD-N0 domain is identified as being solely responsible for the interaction with OutC-HR. The interaction was studied with several techniques. A model of the interacting domains is proposed using the NMR titration result.

5.2 Production of OutD domains

The OutC-HR interaction with OutD occurs in the periplasm. To locate the interacting domains, different OutD N-terminal periplasmic derivatives were produced for interaction studies (Figure 5.1).

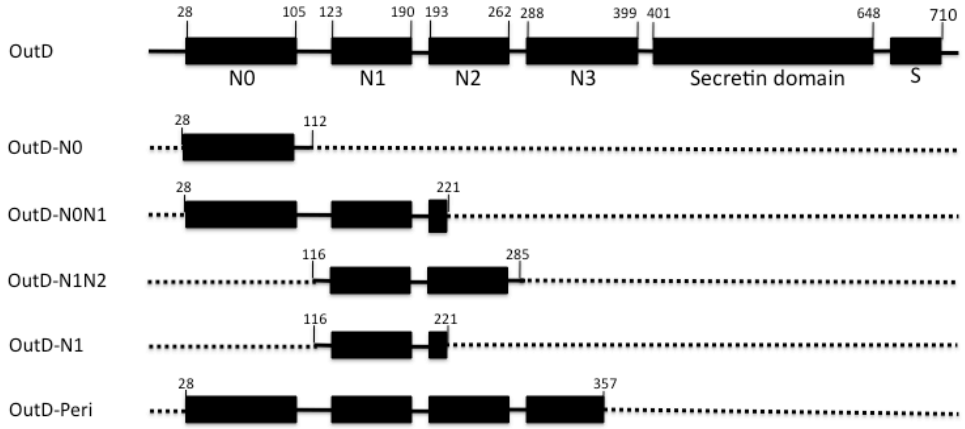


Figure 5.1 Schematic diagram of OutD and its truncated derivatives used in this study. N0, N1, N2 and N3 are the N-terminal periplasmic domains.

5.2.1 Cloning of pET-14b outD-n0

Before receiving pET20b *outD-n0* construct, a pET-14b *outD-n0* construct was originally produced and used for early NMR studies. pET-14b *outD-n0* construct was producing using the pET-20b *outD-peri*, provided by Dr Shevchik (University of Lyon) as the template. The *outD-n0* PCR product was produced with two restriction sites NdeI(5′) and BamHI(3′) and cloned into a pET-14b vector (Figure 5.2). DNA sequencing (MWG) was used to confirm the construct was correct.

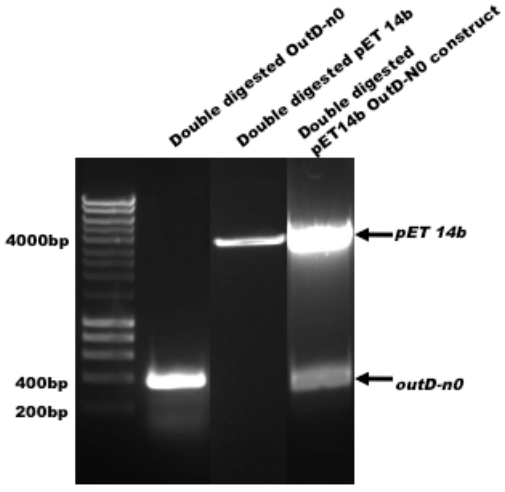


Figure 5.2 Construction of the pET14b-outD-n0.

5.2.2 Expression and purification of OutD derivatives.

OutD constructs in pET20b including *outD-n0, outD-n0n1, outD-n1n2, outD-n1* and *outD-peri* (Figure 5.1),were provided by Dr Shevchik (Lyon). All the plasmids were transformed into *E. coli* BL21 (DE3) cells (see section 2.1.7 for transformation protocols) for protein production. Cells were grown at 30°C until the OD 600 nm reached 0.6 and then grown at 30°C for 3 hours with 1mM IPTG (see section 2.1.8 for conditions and protocol).

Each of these recombinantly produced proteins had an non-cleaveable C-terminal hexahistidine tag and they were initially purified using His Trap HP column (GE Healthcare) (see 2.2.2 for details) and further purified using a size exclusion column. The OutD derivatives all eluted from the size-exclusion column at volumes corresponding to the subunit mass, which indicated the proteins did not aggregate in the conditions used (20mM Tris, pH 7.0 or 5mM NaAc 150mM NaCl pH5.5). This observation supports the view that the periplasmic region of OutD is not responsible for oligomerisation. The purified proteins (Figure 5.3) were subsequently used for interaction studies with OutC-HRF3 by NMR titration.

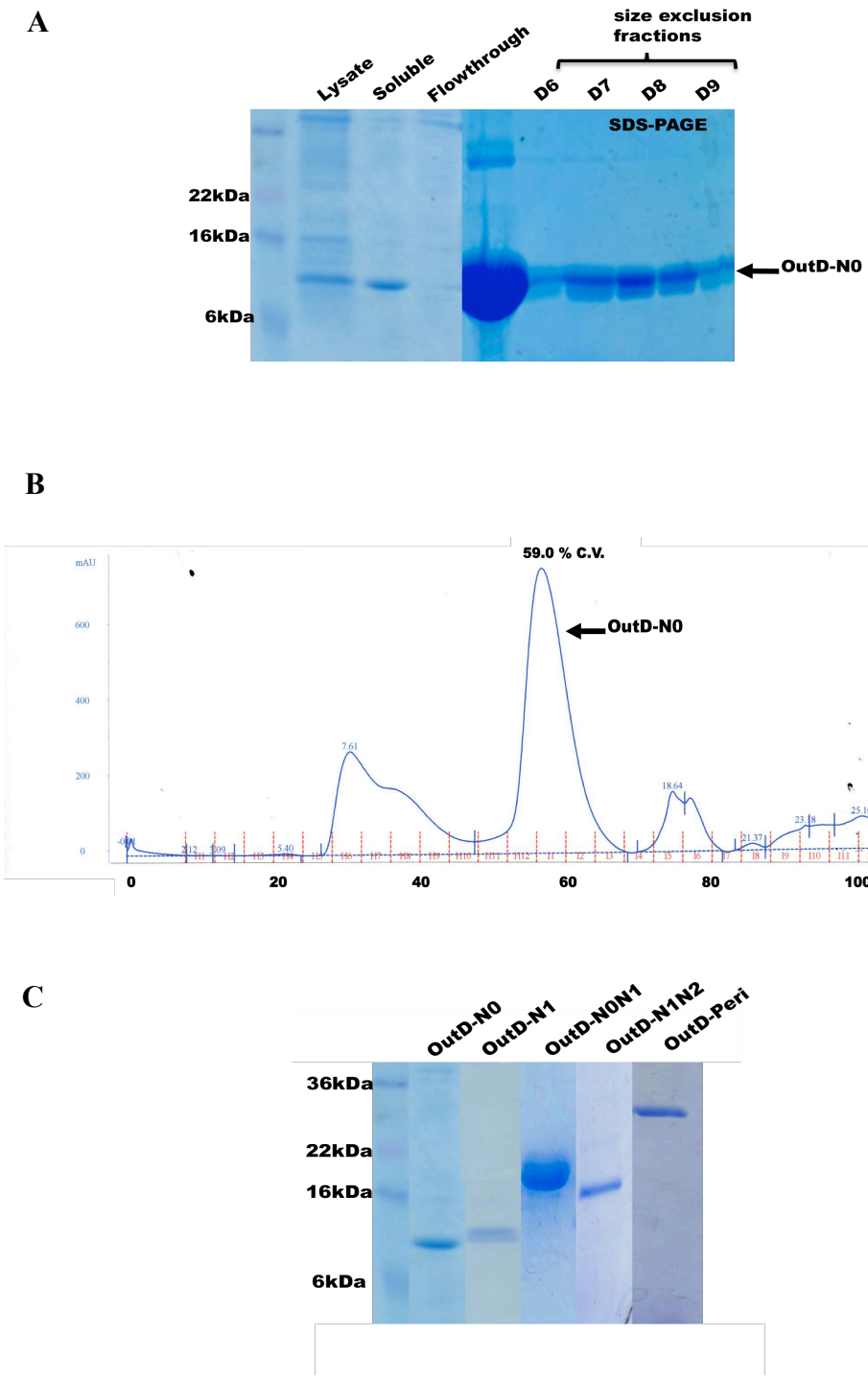


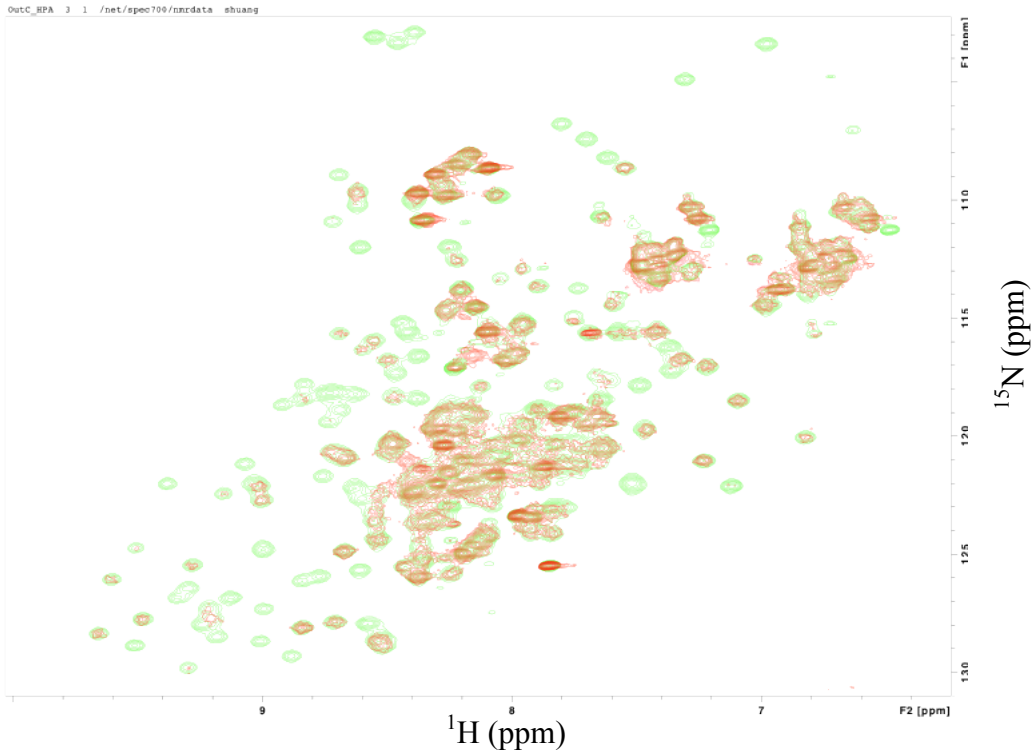
Figure 5.3 Expression and purification of OutD N-terminal proteins (A) Expression and purification of pET-14b OutD-N0. (B) OutD-N0 S200 column size exclusion trace measured at A280. (C) Purified OutD N-terminal proteins. Corresponding proteins obtained after size exclusion chromatograph.

5.3 OutC-HRF3 interacts with OutD-N0

The *outC* and *outD* genes are organized on a single operon (He, Lindeberg et al. 1991) and an interaction between OutC and OutD is known to be important for assembly of a functional T2SS.

Login et al. have shown that in NMR titration using ^{15}N-labeled OutC periplasmic region (OutC-hpa) the spectral peaks corresponding to the PDZ domain do not suffer as much intensity loss as the rest of the protein when unlabeled OutD periplasmic region (OutD-peri) is added (Login, Fries et al. 2010). To investigate if the N0 domain is involved in binding, OutD-N0 was used, in place of the OutD-peri used in the published experiment. OutD-N0 gives very similar effect to that seen with OutD-peri when titrated into labeled OutC-hpa. The ^{1}H-^{15}N HSQC spectrum of OutC-hpa shows some of the peaks loose significant intensity (Figure 5.4), whereas others loose less intensity. The peaks with greatest intensity loss overlay well with the OutC-HRF3 peaks (Figure 5.4B and Figure 5.5). This result supports the view that the OutC-HRF3 and OutC-PDZ domains are arranged in a "beads on a string" manner. The titration confirms that OutC-HRF3 interacts with OutD-N0.

A



B

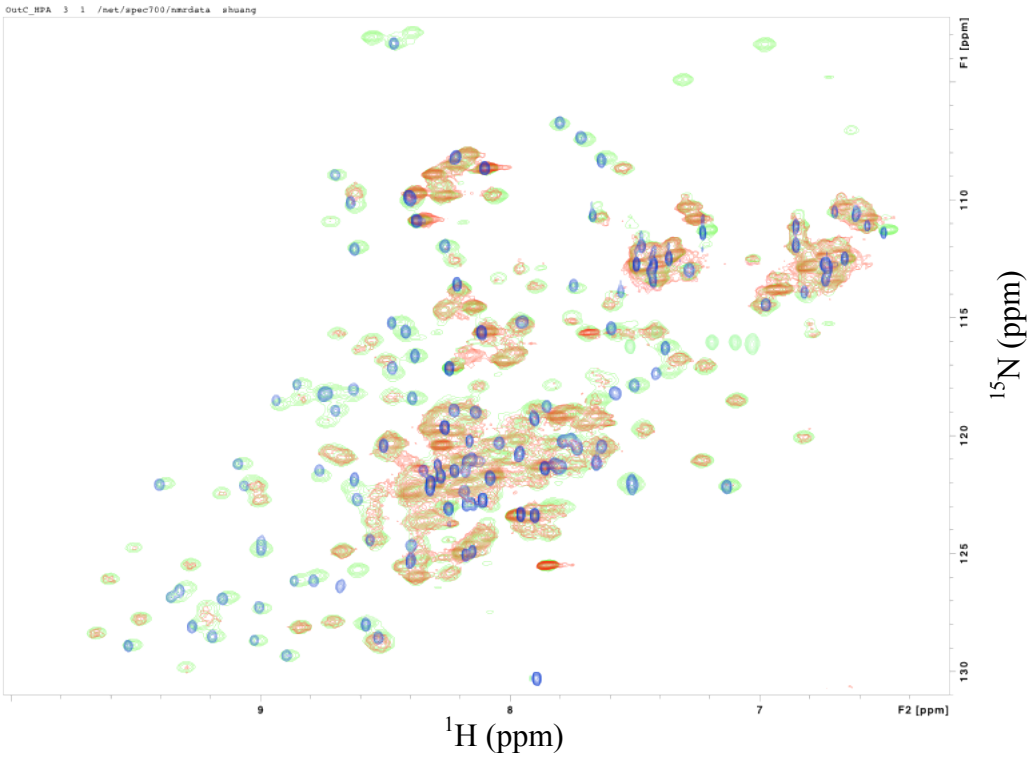


Figure 5.4 OutC-hpa titration with OutD-N0. (A) Overlay of ^{15}N OutC-hpa spectrum (green) with ^{15}N-OutC-hpa with 2 equivalents of OutD-NO (red). (B) Spectra from (A) overlay with OutC-HRF3 spectrum (navy). Spectra acquired at 15°C on Bruker 700MHz in 20mM Tris (pH7.0) and 10% $^{2}H_2O$.

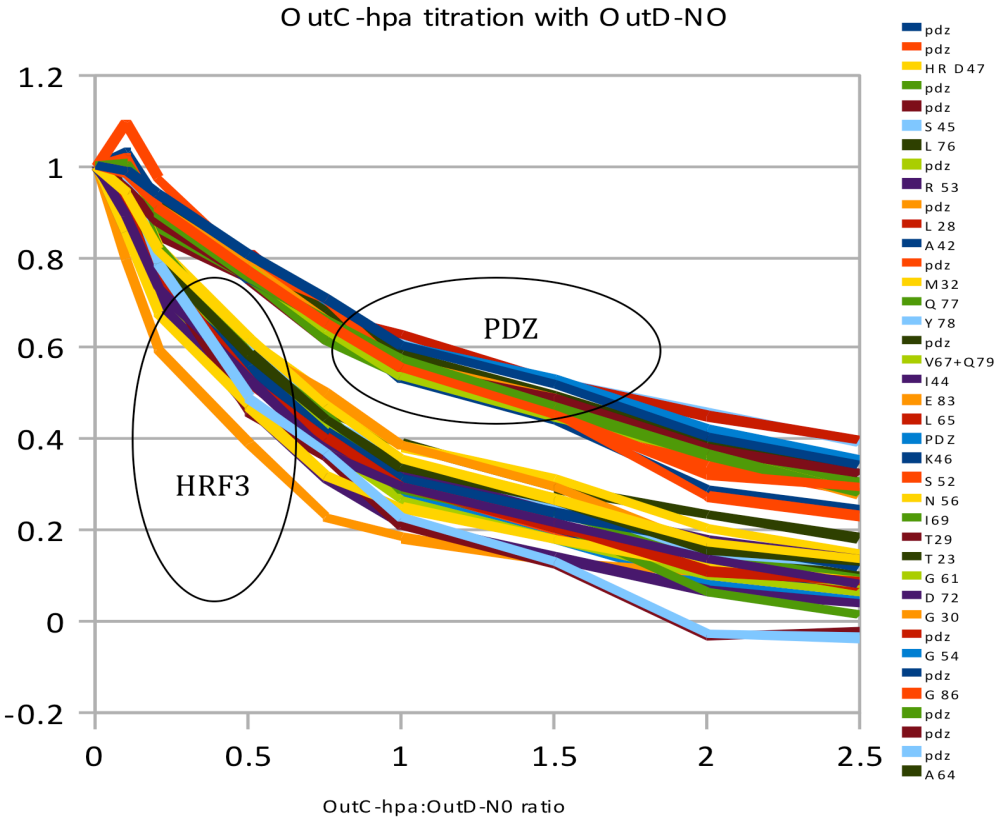
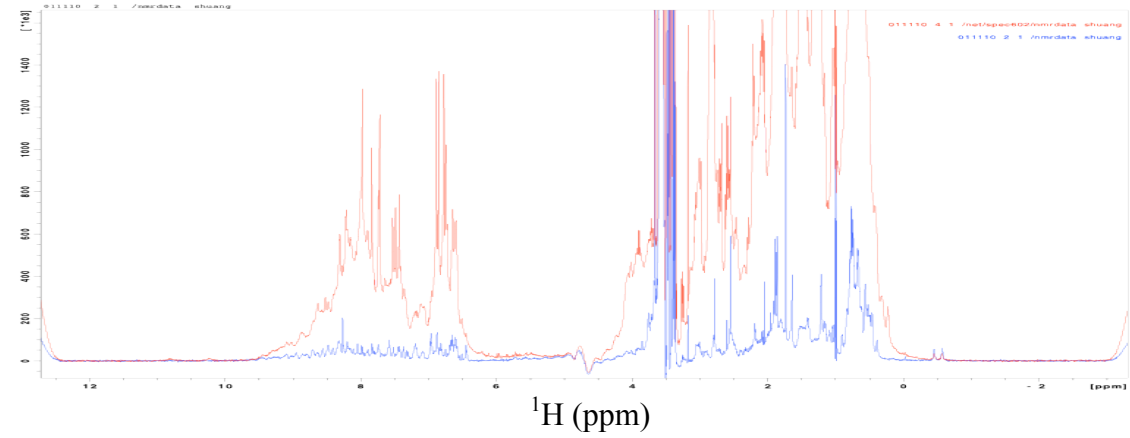


Figure 5.5 The intensity plot of OutC-hpa amide groups during the OutD-N0 titration. OutC-HRF3 residues are labeled.

5.3.1 OutC-HRF3 does not interact N1 or N2 of OutD

To investigate if OutD N1 and N2 are involved in the interaction with OutC-HRF3, OutD-N1N2 was titrated into ^{15}N OutC-HRF3. The 2D-HSQC of labeled OutC-HRF3, with addition of 24 equivalents of OutD-N1N2, showed no obvious peak broadening or peak shifts (Figure 5.6) suggesting there is no interaction with the N1 and N2 domains. However, the possibility of the hinge between N0 and N1 being involved cannot be excluded.

A



B

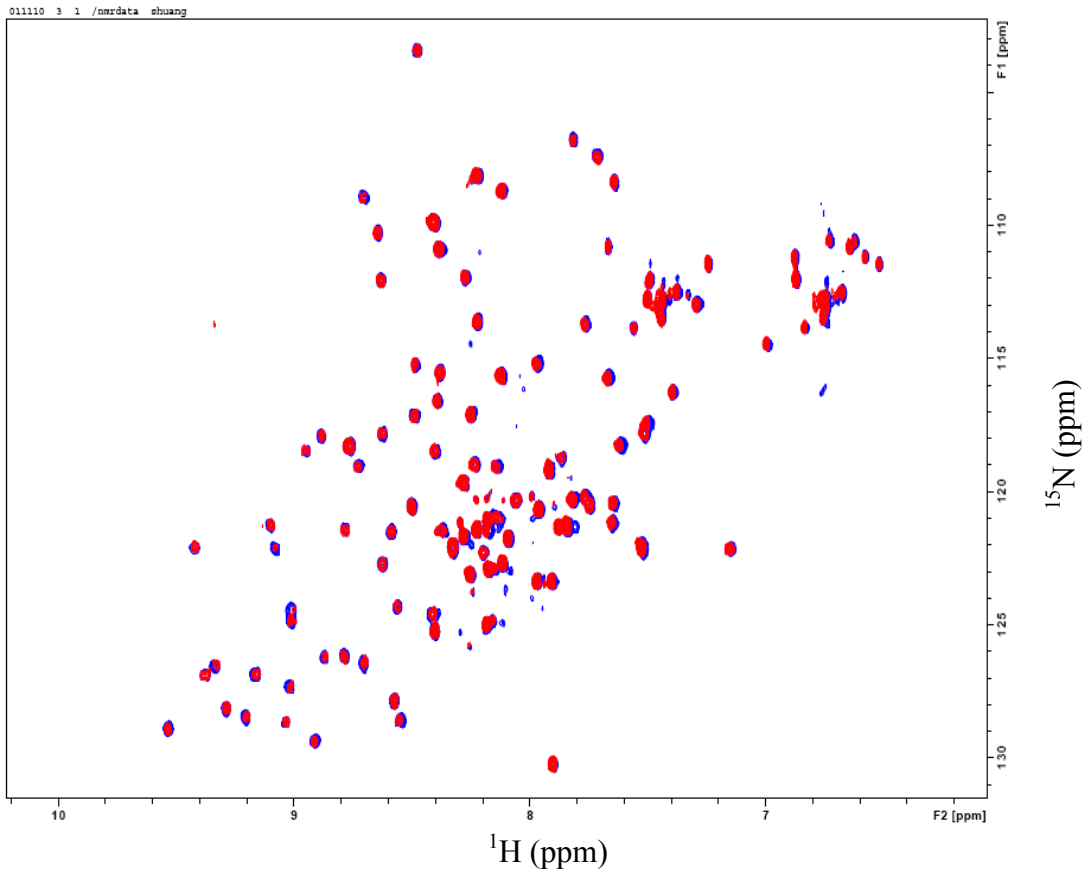


Figure 5.6 ^{15}N OutC-HRF3 titration with OutD-N1N2. Blue spectra, 25mM ^{15}N OutC-HRF3. Red spectra 25mM ^{15}N OutC-HRF3 titrated with 600mM OutD-N1N2. (A) Overlay of 1D spectrum. (B) Overlay of HSQC spectra.

5.4 NMR study of OutD-N0

In order to understand the interaction between OutC-HR and OutD N-terminal domain at the molecular level, different OutD constructs, corresponding to different N-terminal subdomain(s) were produced, expressed, purified and tested against the OutC-HRF3. The OutD-N0 domain was found to interact with OutC-HR. Therefore a combination of NMR studies and structure modulation were employed using OutD-N0. The OutD and OutD-N0 structures are modulated according to the known GspD structure. The OutD-N0 NMR structural studies were continued to pave the way to understand the interaction at the molecular level.

5.4.1 Backbone assignment of OutD-N0

The backbone assignment for 65 out of 80 residues of OutD-N0 was achieved (see section 3.1.2 for methodology; Figure 5.7). The assignment was poor for residues located on β1 and β3 (Figure 5.19) because these residues showed very weak intensities and can hardly be observed in 3D spectra. Further experiments with higher field strengh spectrometer and longer aquisation time could help with assignment of these missing residues. However, as α1 and α2 were well assigned, for N0. These two helices was used to alocated and confirm interaction surface.

5.4.2 OutD-N0 secondary structure prediction

The OutD-N0 backbone chemical shifts were used to predict the secondary structure using TALOS+. The predicted secondary structure is consistent with the available GspD-N0 secondary structure (Figure 5.8), consisting of two α-helixes and 5 short β-strands. The predicted secondary structure is consistent with the known GspD-N0 secondary structure (Figure 5.8B).

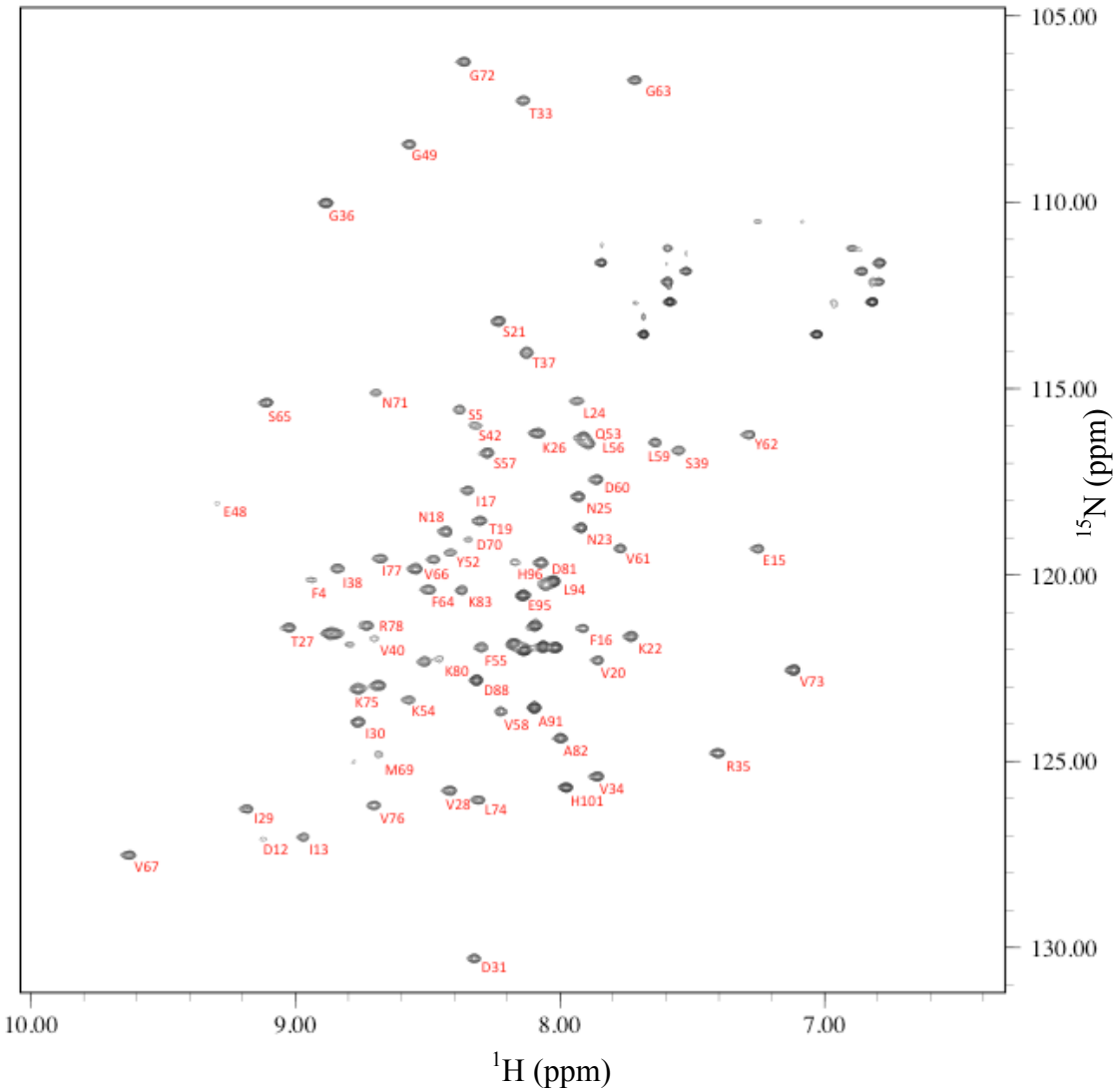


Figure 5.7 ^{1}H-^{15}N HSQC spectrum of OutD-N0. The spectrum was acquired at 600MHz at 15°C in 20mM Tris (pH 7.0) and 10% $^{2}H_2O$.

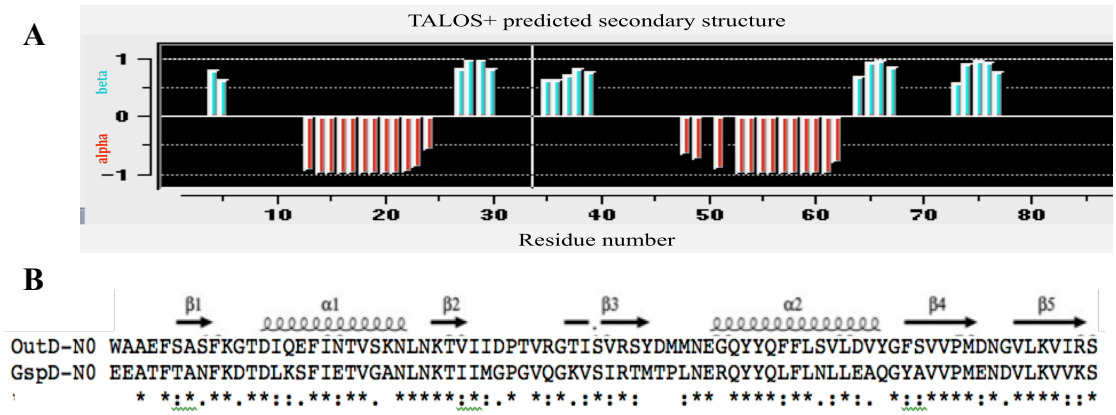


Figure 5.8 Secondary structure prediction of the OutD-N0. (A) TALOS+ predicted OutD-N0 secondary structure. (B) Sequence and secondary structure alignment of OutD-N0 and the only known structure GspD (PDB code: 3EZJ from *Vibio cholera*).

5.4.3 Relaxation study

The protein backbone dynamics of OutD-N0 were investigated using ^{15}N relaxation measurements. T_1, T_2 and ^{1}H-^{15}N-NOE relaxation experiments were performed to understand internal dynamics at different frequencies and aspects.

The T_1, T_2 and ^{1}H-^{15}N-NOE relaxation profiles for OutD-N0 are shown in Figure 5.9. In the relaxation profile, missing residues correspond to proline residues, overlapped peaks or peaks with extremely weak intensity. The OutD-N0 domain is a well defined structure with T_1 and T_2 stable throughout the sequence. The N-terminal showed higher T2 and lower NOE than average, indicate the flexibility at the N-terminus.

Using the average T_1 and T_2 value from the well structured region, from residue 15 to 78, the OutD-N0 correlation time was calculated to be 9.37 ns. For a 10kDa protein, the correlation time of 9.37 ns, indicates the protein is well structured and behaves as a subunit in solution. However, the OutD-N0 correlation time is slightly longer than OutC-HRF3(11 kDa, 8.59 ns) and EspC-HR (11kDa, 8.68), all of similar molecular weight, this likely reflects the shape of the different domains.

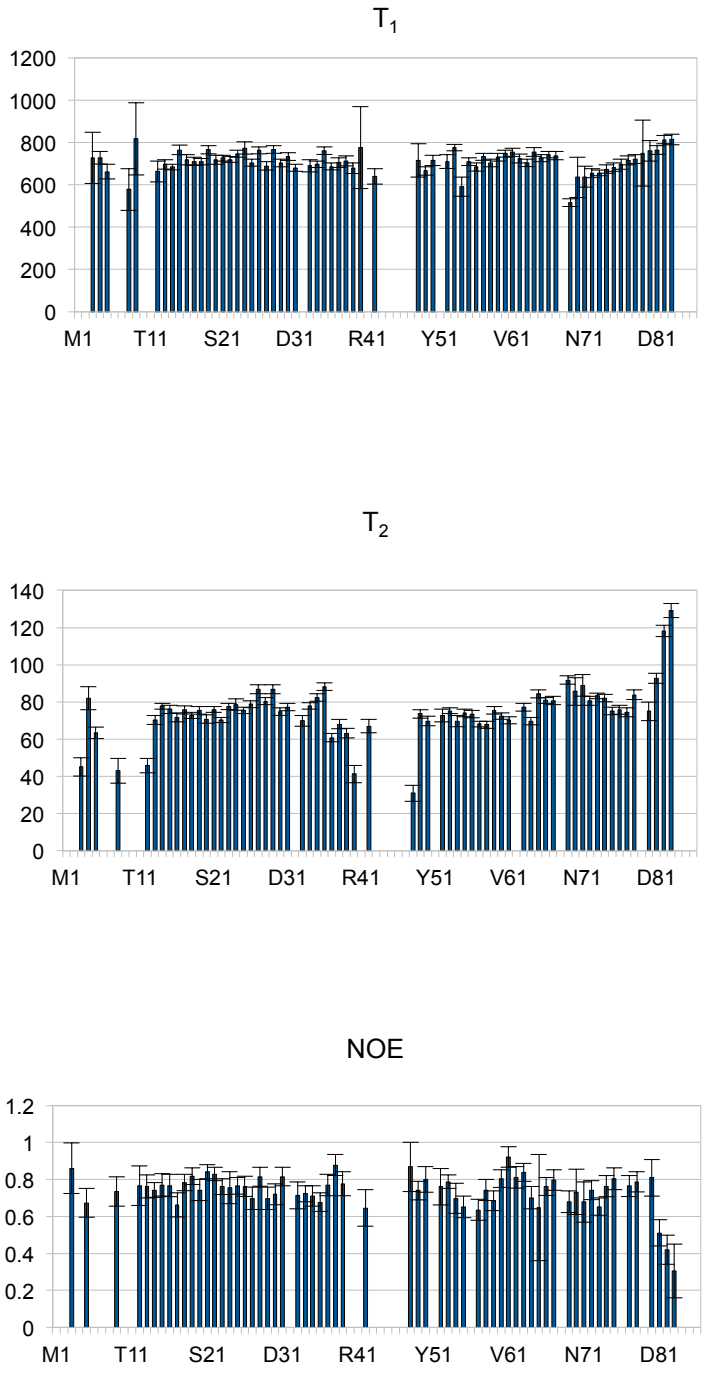


Figure 5.9 OutD-N0 T_1,T_2 and NOE relaxation, acquired at 15°C on Varian 600 MHz.

5.4.4 OutD-N0 structure modulation

The OutD-N0 chemical shift based secondary prediction is consistent with the known GspD-N0 structure region. Sequence identity between OutD-N0 and GspD-N0 is 52%, This indicates that both N0 domain share similar fold. Therefore SWISS-MODEL (http://swissmodel.expasy.org/) and 3D-JIGSAW (http://bmm.cancerresearchuk.org/~3djigsaw/) were used to generate an OutD-N0 model according to sequence alignment and homology with the GspD-N0 structure (PDB:3EZJ). The structures of OutD-N0 generated by two different servers are based on the same known structure. Therefore are very similar with backbone RMSD of 0.25**Å** (Figure 5.10). The generated models give some idea of the molecule backbone orientation. The sidechain arrangement are not necessarily represented. .

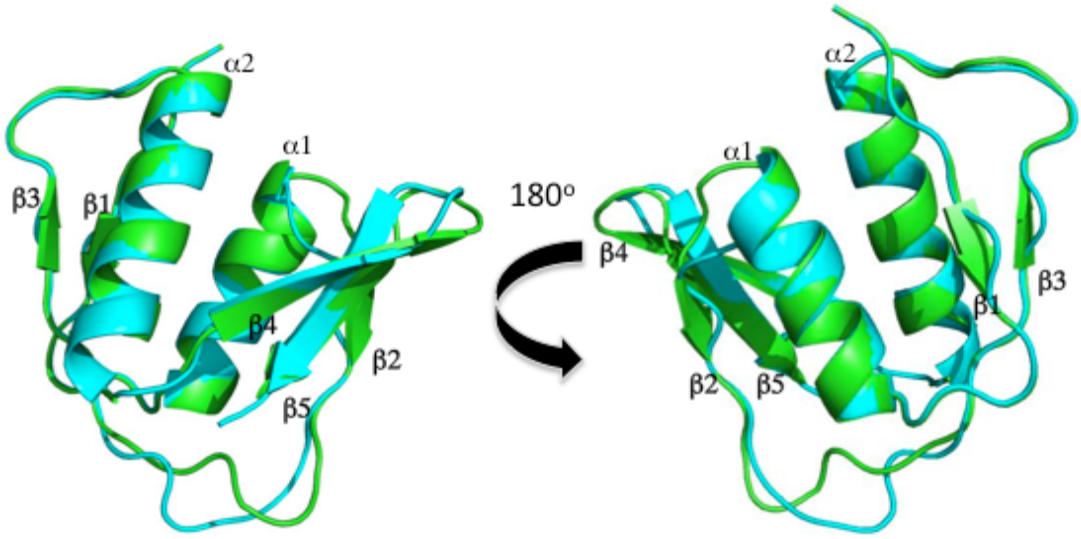


Figure 5.10 Homology based OutD-N0 structure prediction. SWISS-MODEL generated structure in green. 3D-JIGSAW generated structure in cyan.

5.4.5 GspD N0 and N1 domain arrangement

The periplasmic region of the GspD structure has been solved in the presence of nanobody (Korotkov, Pardon et al. 2009). The nanobody binds to both N0 and N1 (Figure 5.11), sharing a total buried surface 1896 Å. This contact is large compared to

the N0 and N1 contact surface of 1180Å. Therefore the N0, N1 domain orientation in the crystal structure is not necessary reflecting the *in vivo* domain orientation.

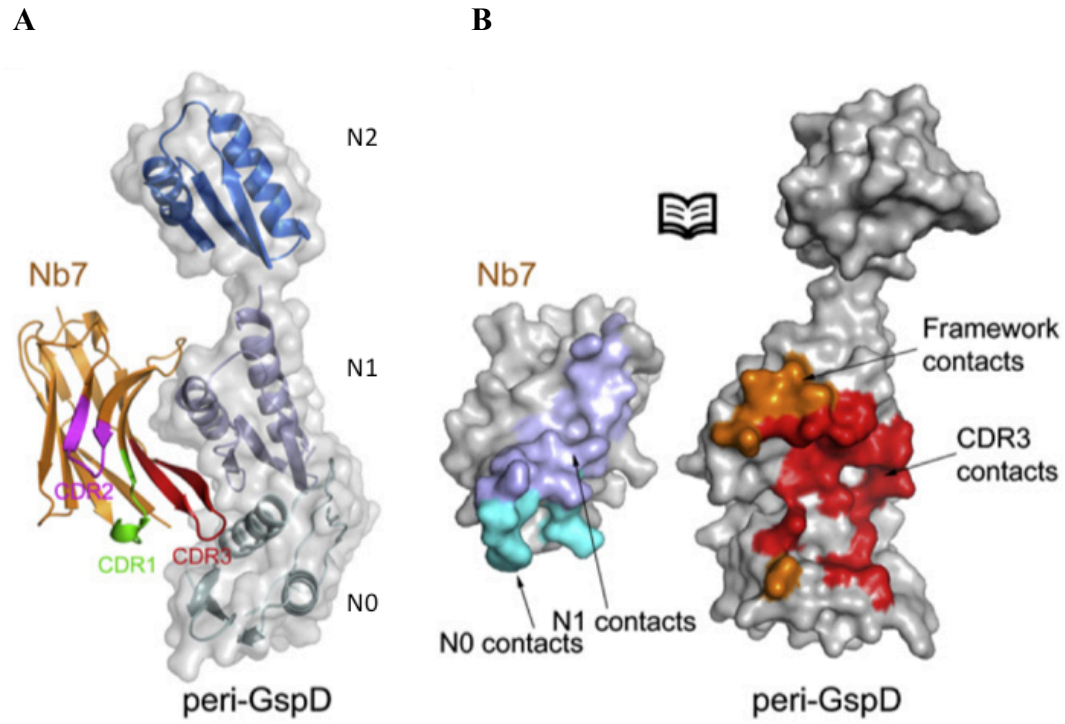


Figure 5.11 Interaction between Peri-GspD and the nanobody. (A) Structure of the peri-GspD:nanobody dimer with nanobody in orange, CDR1 in green, CDR2 in purple, and CDR3 in red, with the peri-GspD subdomains (B) An ‘‘open book’’ representation of the peri-GspD:Nb7 interface with foot- prints in colors according to interacting partner (CDR3 footprint, red; framework, orange; N0 subdomain, cyan; N1 subdomain, light blue). Figure adapted from (Korotkov, Krumm et al. 2006)

To confirm the interaction persists in solution NMR titration was used to examine the interaction between OutD-N0 and OutD-N1N2 (Figure 5.12). When OutD-N1N2 was titrated into ^{15}N-labeled OutD-N0, the signal intensity of OutD-N0 became weaker with addition of more OutD-N1N2. The signal loss behaves in a fast exchange manner, suggesting that the interaction is weak.

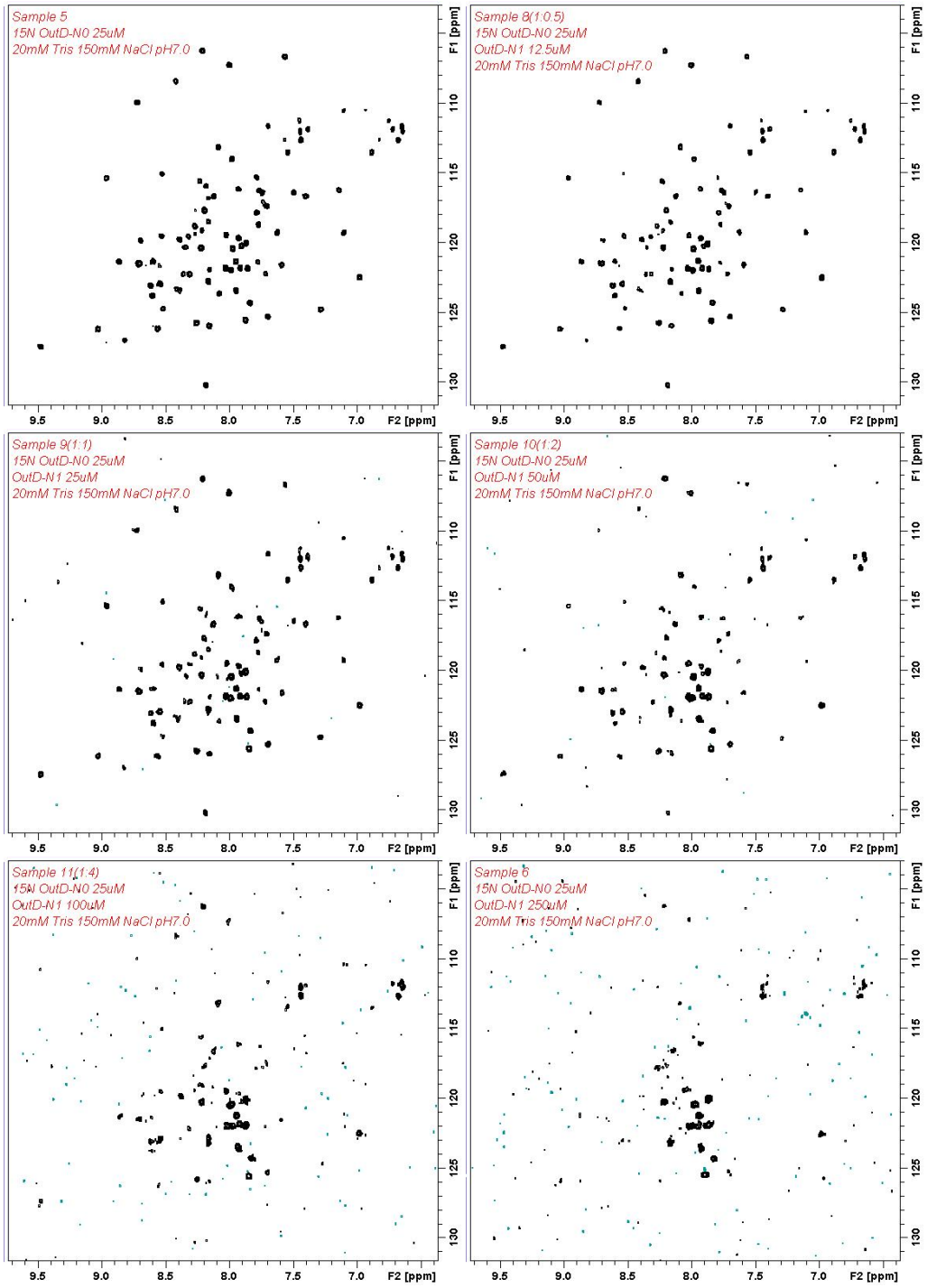


Figure 5.12 ^{1}H-^{15}N HSQC of OutD-N0 in the absence and the presence of OutD-N1N2. ^{15}N OutD-N0 was used at 25μM, OutD-N1N2 concentration as indicated. Titrations were in 20mM Tris, 150mM NaCl, pH7.0 using Bruker 600 at 15^{O}C.

To investigate the solution orientation of N0 and N1 domains the OutD-N0 and OutD-N0N1 ^{1}H-^{15}N HSQC were measured separately at 15℃ in 20mM Tris pH 7.0 (Figure 5.13). From the spectra, peaks belong to N0 domain can be recognized on the OutD28-221 spectrum. However there is slight peak shifting when comparing the peaks belonging to N0 on the OutD-N0N1 spectrum and OutD-N0. The peak shifts are likely induced by the presence of the N1 domain next to N0. To locate the contact surface on N0 according to chemical shift changes, chemical shift variations (Δδav) due to presence of the N1 domain were calculated according to (Pellecchia, Sebbel et al. 1999; Korotkov, Pardon et al. 2009):

$$\Delta\delta_{av}=\{0.5[\Delta\delta\ (^{1}H)^{2}+(0.2\Delta\delta\ (^{15}N))^{2}]\}^{1/2}$$

where $\Delta\delta_{av}$ is the combined chemical shift perturbation and ^{1}H and ^{15}N are the proton and nitrogen chemical shifts in parts per million (ppm) for the backbone amide of each residue.

OutD-N0 peak and the closest OutD-N0N1 peak were used to calculate Δδav for individual residue. Plotting the residues in OutD-N0 domain with $\Delta\delta_{av}$ greater than 1 ppm, a patch is revealed (Figure 5.14). In the presence of the N1 domain, the patch has the most chemical shift changes are likely to locate close to the interface between N0 and N1. N0 showed greater chemical shift changes are considered to be at or closer to the N1 domain. However, this patch does not locate at the N0 N1 interface within the crystal structure in the presence of nanobody. Therefore the interaction between N0 and N1 in GspD N-terminal crystal structure could be induced by presence of nanobody or crystal packing. The assignment of Out-N0 was not complete on β1 and β3 region. However, the α1 and α2 region is relatively complete. If we do expect the N0 N1 contact observed in crystal structure, residues located on α2 should have given

greater $\Delta\delta_{av}$ than the other residues.

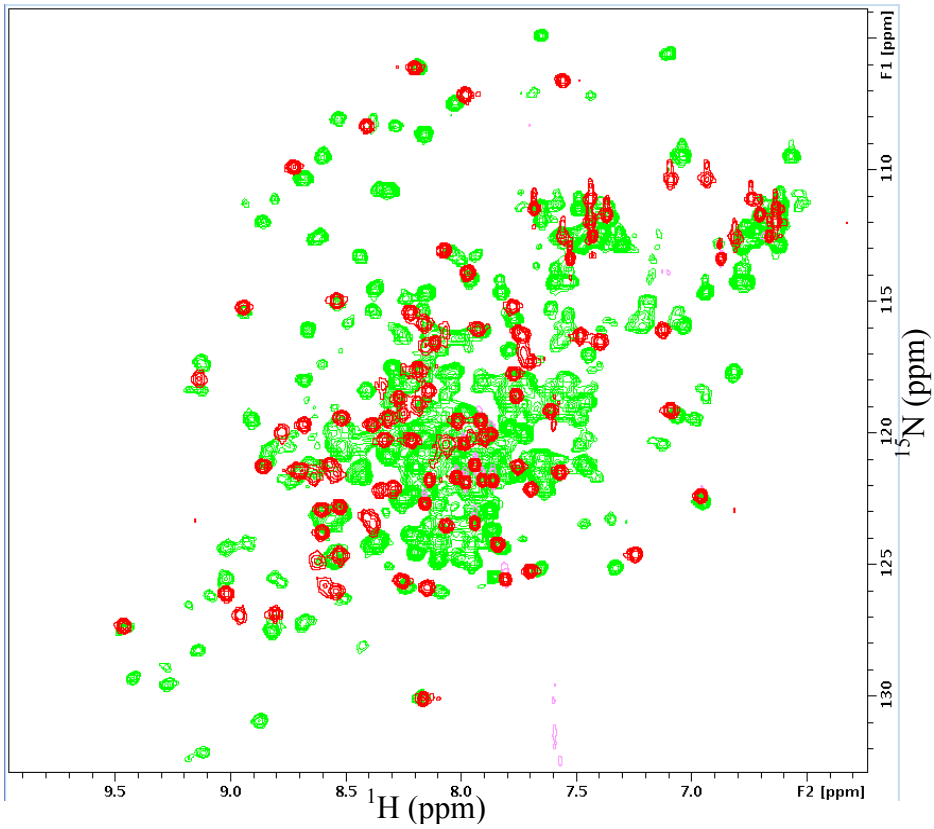


Figure 5.13 Overlay of ^{1}H-^{15}N HSQC spectrum of OutD-N0 (red) and OutD-N0N1(green) spectrum. Both spectra were acquired at Bruker 600MHz at 15°C in 20mM Tris (pH 7.0) and 10% $^{2}H_2O$.

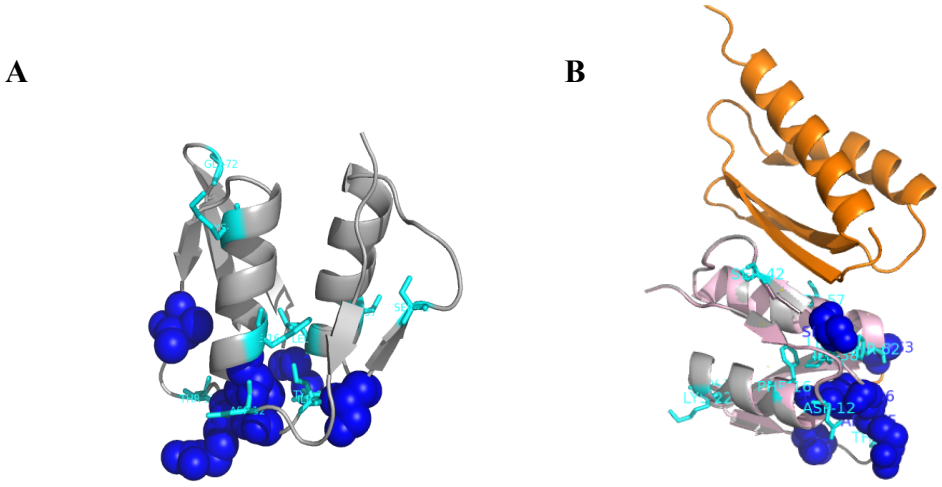


Figure 5.14 $\Delta\delta_{av}$ is a weighted average of the ^{15}N and $^{1}H^{N}$ chemical shift. The easily recognized N0 peaks were paired and $\Delta\delta_{av}$ are calculated. The residues with $\Delta\delta_{av}$>1ppm are shown in navy as spheres and residues with 0.5 < $\Delta\delta_{av}$ < 1 ppm are shown in cyan as sticks. OutD-N1 domain (in orange) was superimposed to the OutD-N0 according to the orientation seen in the homologous (PDB: 3EZJ) crystal structure.

5.5 Interaction study of OutC-HRF3 and OutD-N0 by NMR

To identify the interaction surface between OutC-HRF3 and OutD-N0, a series of titrations between these two proteins were made using the changes in the ^{1}H-^{15}N HSQC spectra.

When titrating OutD-N0 into ^{15}N OutC-HR3 or vice versa, a loss of intensity of all the peaks was observed (Figure 5.15 and Figure 5.16). There was no dilution between titration points because the OuC-HRF3 concentration was maintained as described Chapter 3.7. The peak weakening effect is most likely caused by the increase in the size of the proteins after complexation, which slow down the tumbling of the molecule.

The diminution of peak intensitiesare accompanied with peak shifts (Fig. 6.13) in the titration experiments. Interaction between OutC-HRF3 and OutD-N0 appeared to be weak, as addition of 10 equivalents of OutD-N0 did not saturate the interaction and the peak weakening effect was still observable (Figure 5.15). However, the titration results show the interaction behaves consistently when titrating OutC-HRF3 in OutD-N0 or the other way round (Figure 5.15 and Figure 5.16). The peak shifts indicate the proteins are behaving in a fast exchange manner.

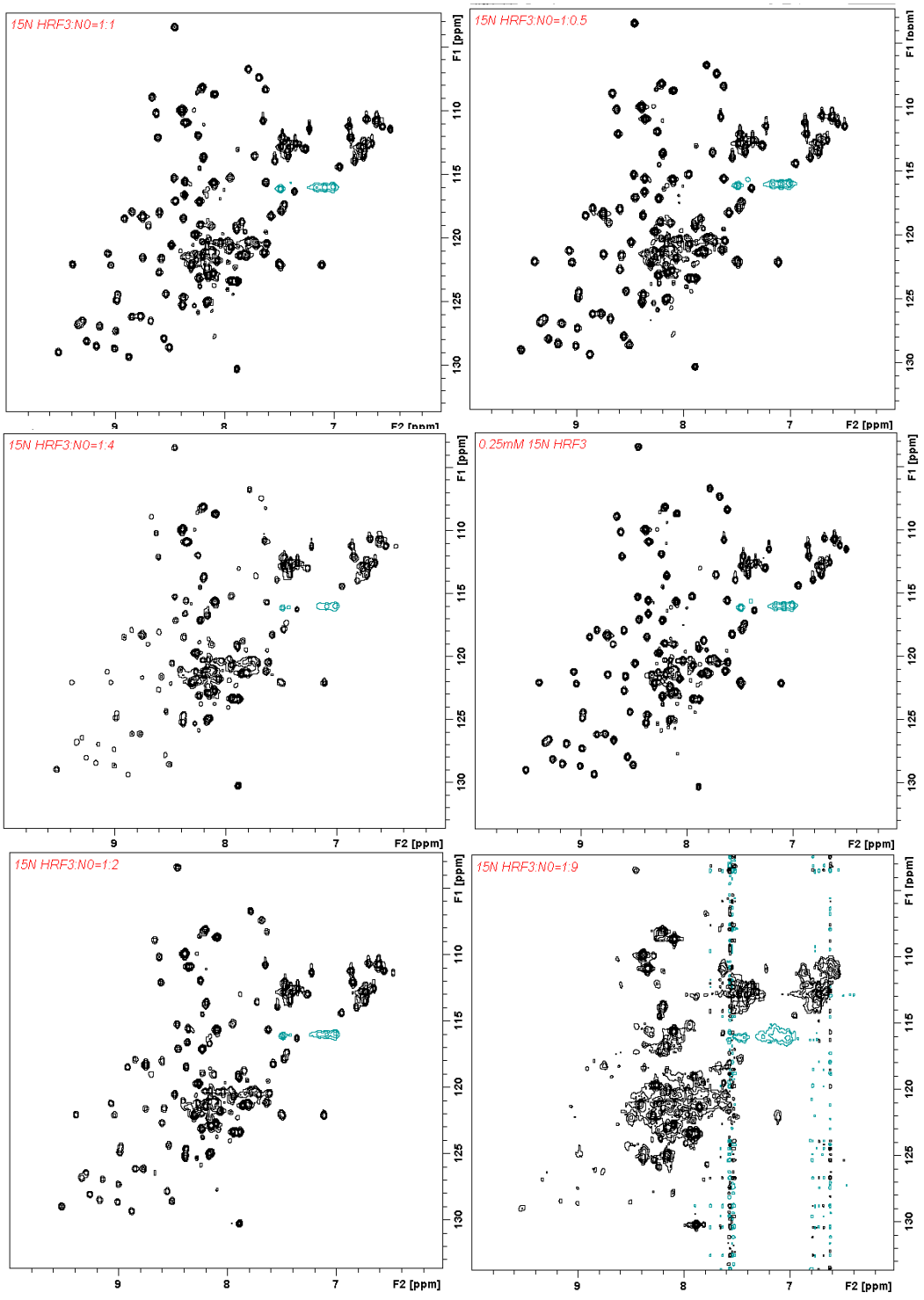


Figure 5.15 ^{15}N HSQC of OutC-HRF3 in the absence and the presence of OutD-N0. OutC-HRF3 was used at 250 µM.

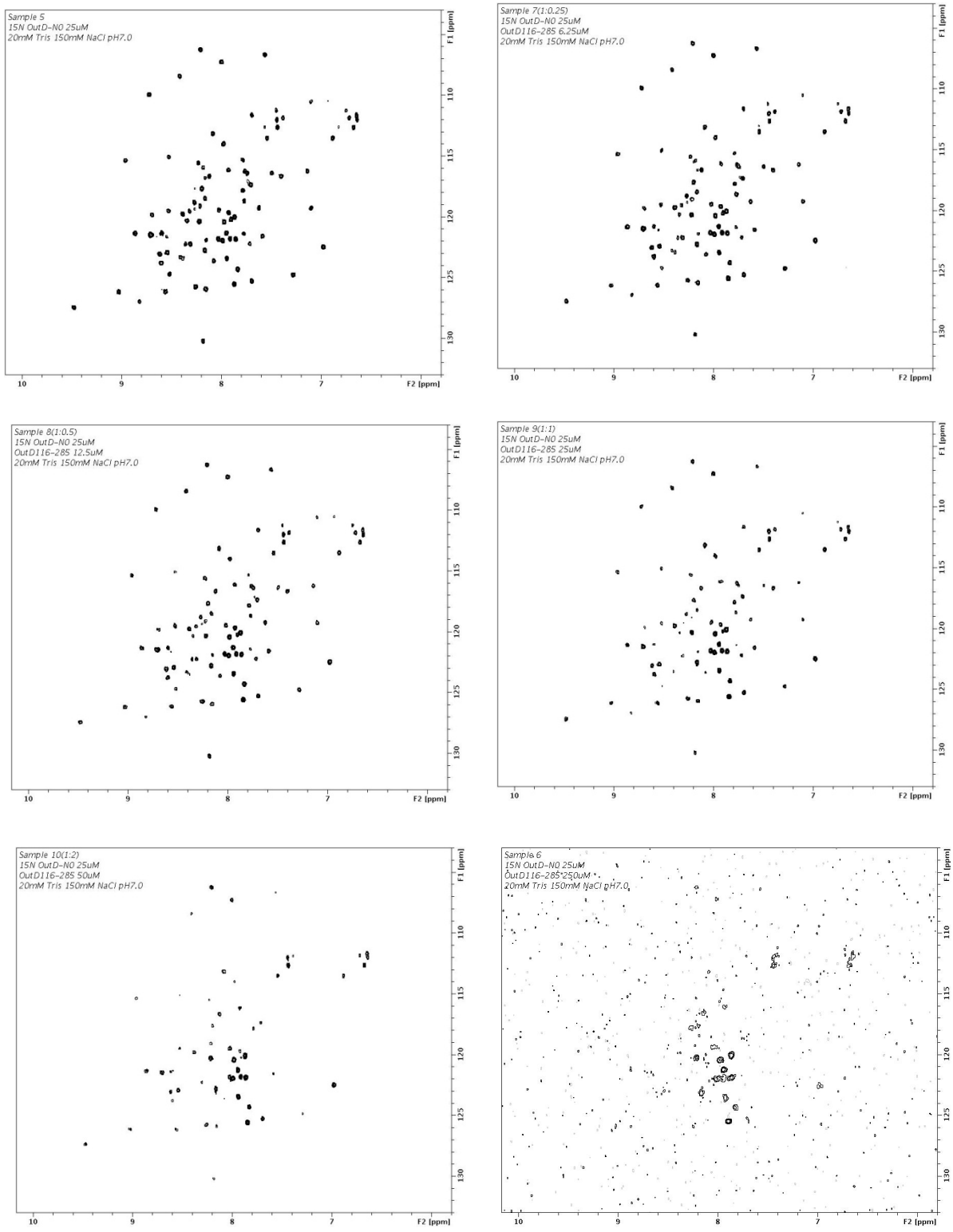


Figure 5.16 ^{1}H-^{15}N HSQC of OutD-N0 in the absence and the presence of OutC-HRF3. OutD-N0 was used at 25 μM.

To check if the interaction between OutC-HRF3 and OutD-N0 is disturbed by the presence of OutD-N1N2, OutD-N1N2 was titrated into ^{15}N OutC-HRF3 and OutD-N0 mixture (Figure 5.17). With addition of 8 equivalents of OutD-N0, the OutC-HRF3 HSQC signal is weakened. Then with the further addition of 270μM of OutD-N1N2, which is over ten times higher than the 25μM OutC-HRF3 and more than one equivalent of OutD-N0 concentration, no obvious changes were observed. From this experiment, if OutD-N1N2 interacts OutD-N0 in similar manner as OutC-HRF3, the significantly higher concentration OutD-N1N2 would have displaced OutC-HRF3 and restore spectra signal loss. Hence, I can conclude that OutD-N1N2 does not compete with OutC-HRF3 for OutD-N0 binding.

Both OutC-HRF3 and OutD-N0 HSQC showed some peaks shifting when titrating with the other (Figure 5.18). This indicates that there are some local structural changes when OutC-HRF3 interacts with OutD-N0. Mapping the peaks shifted in the titration on the NMR OutC-HRF3 structure and computer generated OutD-N0 model, allows the interaction surfaces to be identified (Figure 5.19).

The residues that show chemical shift changes are mainly located on β1 of OutC-HRF3 and α2 and β3 of OutD-N0 (Figure 5.19). The interaction surfaces on both molecules comprise hydrophobic patches. Hence it is vealy like that OutC-HRF3/ OutD-N0 interaction is driven by hydrophobic interactions. To use the interaction surfaces from both proteins and assemble the complex, further experiments needed to be done to confirm the orientation of both proteins in the complex.

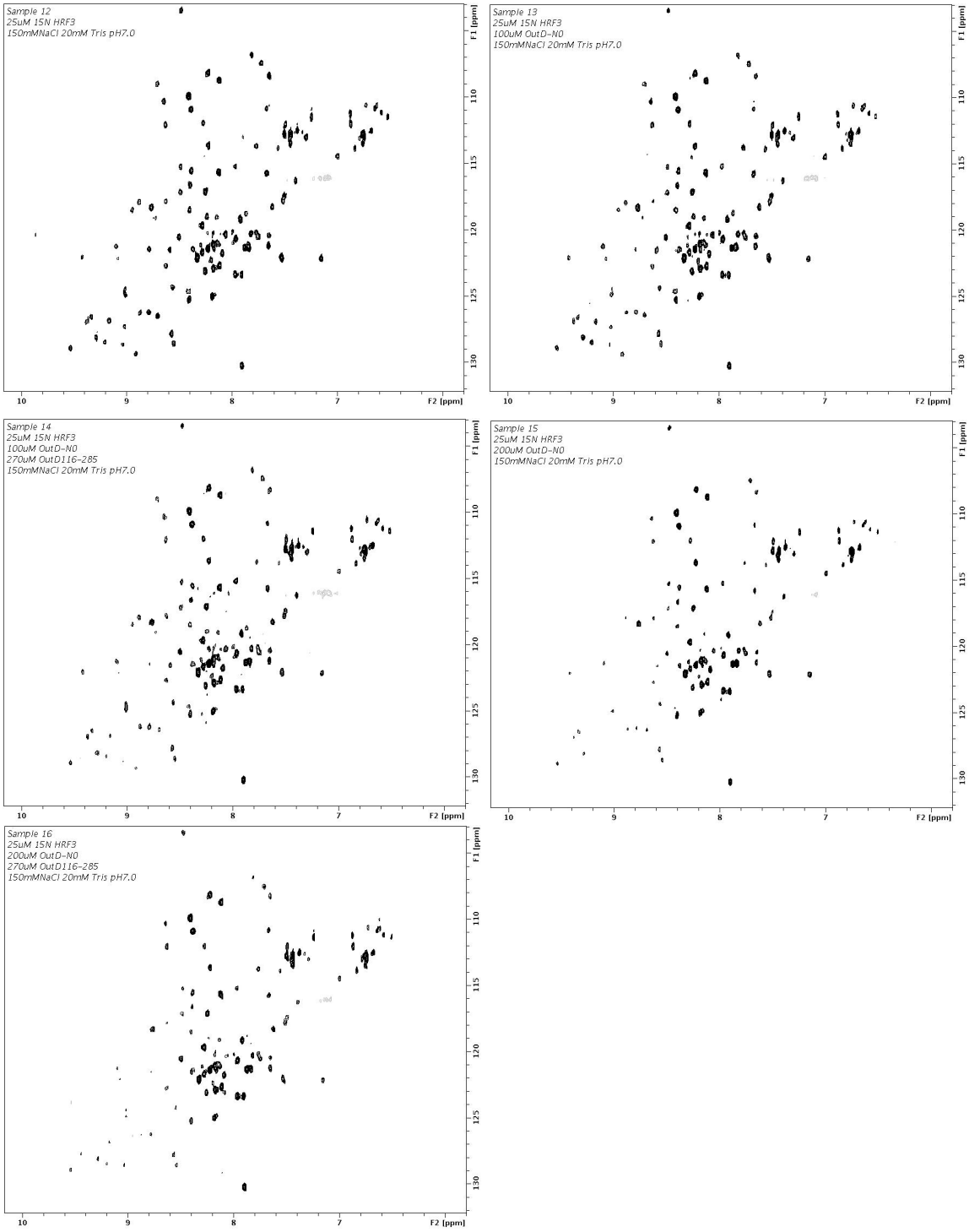
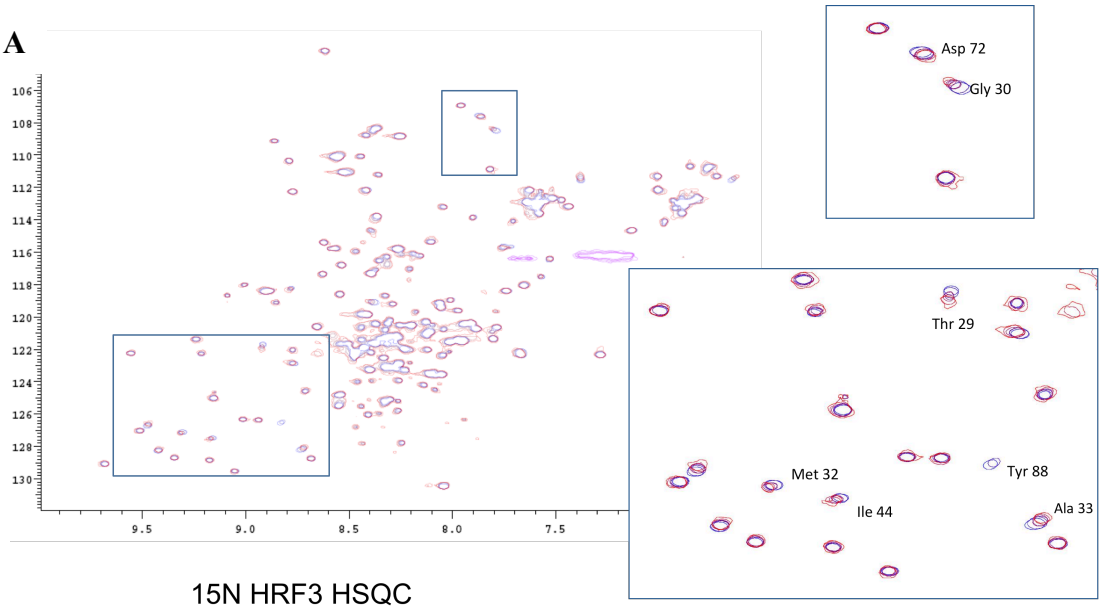


Figure 5.17 Effect of OutD-N1N2 on the OutC-HRF3 and OutD-N0 mixture. The effect was monitored by recording ^{15}N HSQC of OutC-HRF3. ^{15}N OutC-HRF3 was measured at 25μM. The sample composition was as labeled. All sample prepared in 20mM Tris 50mM NaCl pH7.0.



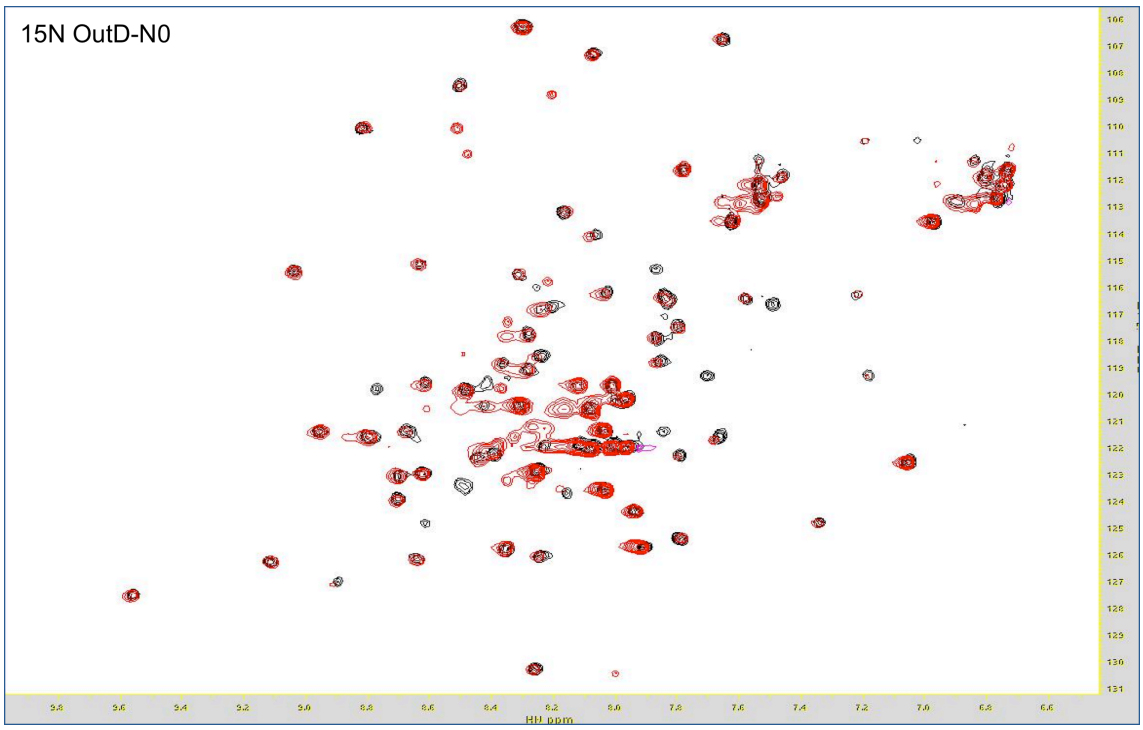


Figure 5.18 ^{15}N labeled OutC-HRF3 titrated with non-labeled OutD-N0. Overlay of six ^{1}H-^{15}N HSQC spectra of OutC-HRF3 with ratio to OutD-N0 of 1:0 (Blue), 1:0.2 (Navy), 1:0.4 (Purple), 1:0.6 (Magenta), 1:0.8 (Pink), 1:1 (red). OutC77-172, 0.2mM. B ^{15}N labeled OutD-N0 titrated with OutC-HRF3. Overlay of two ^{15}N OutD-No HSQC with ratio of ^{15}N OutD-N0 to OutC-HRF3 of 1:0 (black) and 1:2(red) ^{15}N OutD-N0 ,0.2mM.

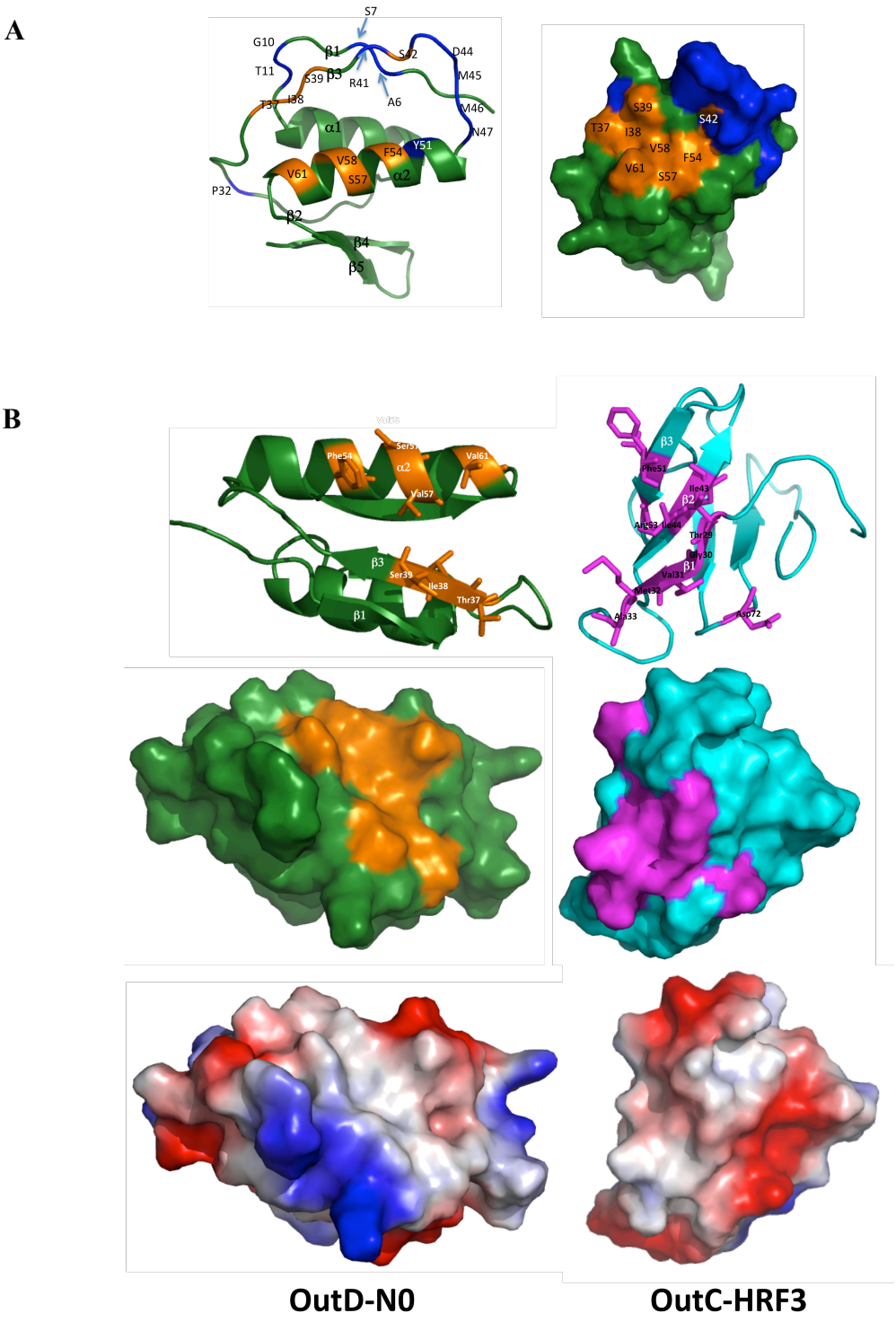


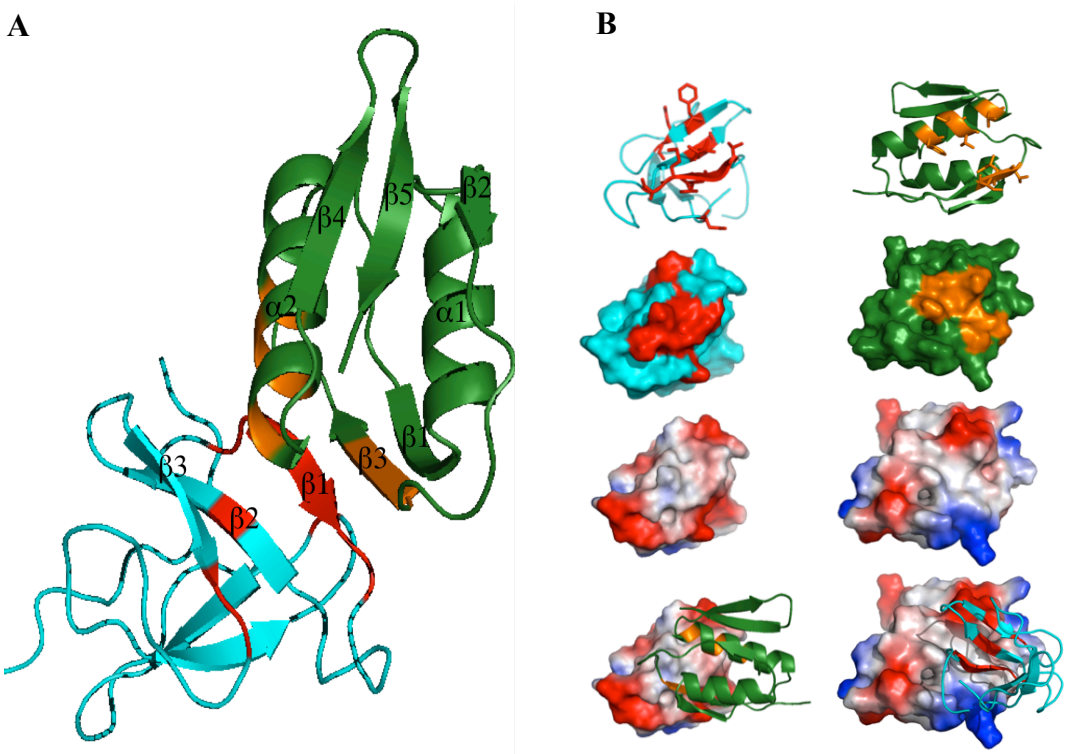
Figure 5.19 Interaction surface identified by chemical shift perturbation (A) OutD-N0 interaction surface in orange with unassigned residues in blue. (red). (B) OutD (orange) and OutC-HRF3 (magenta) interaction surface in cartoon, surface and electrostatics plots (shown in arbitrary PyMOL unit).

5.6 Model of OutC-HRF3 and OutD-N0 interaction by HADDOCK

Haddock was use to dock OutC-HRF3 and OutD-N0 according to the titration result (Dominguez, Boelens et al. 2003). The residues involved in the interaction were introduced as Ambiguous Interaction Restraints (AIRs) to drive the docking process.

The OutD-N0 interaction region is located between α2 and β3. The OutC-HRF3 interaction region is located at β1 strand. HADDOCK docks the OutC-HRF3 β1 strand between α2 and β3 of OutD-N0, which leads to β1, β3 from OutD-N0 and β1, β2, β3 from OutC-HRF3 forming a continuous antiparallel β-sheet. The complex is mainly stabilized by hydrophobic interactions, involving hydrophobic patches formed byV31 M32 A33 of OutC-HRF3 and I38 F54 V58 V61of OutD-N0.

Looking into GspD-N0 structure (Figure 5.21), it consists two hydrophobic surfaces, one between α2 and β3 the other one between α1 and β2. Here the NMR titration result suggest that OutC-HRF3 docks on the hydrophobic surface between α2 and β3. However, in recent GspC-GspD co-crystallisation structure, The hydrophobic path between α2 and β3 of GspD is occupied by the GspD-N1 domain, where GspC-HR is docking on to the hydrophobic path between α1 and β1(Korotkov, Johnson et al. 2011). It is intresting to note that the interactions between both GspD-N0/GspC-HR and GspD-N0/GspD-N1 are all mainly coordinated by hydrophobic interactions, with no salt bridges or hydrogen bonds. With hydrophobic interactions the partner could be easily swapped during conformational changes in the assembled secretion system.



C

```
OutC-HRF3          10        20        30        40        50        60        70        80
E.chrysanthemi     AGALDASQMSNLPPSTLNLSLTGVMAGDDDSRSIAIISKDNEQFSRGVNEEVPGYNAKIVSIRPDRVVLQYQGRYEVLGL
E.carotovora       SAVNDAALSGDIPLSSLNISLTGVLASEDAKRSIAIIAKDSQQYSRNVGDAVPGYEAKIVTISADRVVLQYQGRYEALHL
K.pneumoniae       PAPTDATHLNQVPVSSLKLRLTGLLASSNPARSIAIIEKGNQQVSLSSGDTIPGYDARIVAILPDRIIVNYQGRKEAILL
V.cholerae         APVVEQPVVVDAPKTRLSLVLSGVVASNDAQKSLAVIANRGVQATYGINEVIEGTQAKLKAVMPDRVIISNSGRDETLML
E.coli             RAQETMPSLSNDLLSGEDLDVRGILYSSVAEHSVAIFAHNNRQFSLSVGEKVPSYDATISAIFSQHIVINYQGKTVSLPL
                                     β1        β2        β3        β4        β5        β6

OutD-N0            10        20        30        40        50        60        70
E.chrysanthemi     EFSASFKGTDIQEFINTVSKNLNKTVIIDPTVRGTISVRSYDMMNEGQYYQFFLSVLDVYGFSVVPMDNGVLKVIR
E.carotovorum      EFSASFKGTDIQEFINTVSKNLNKTVIIDPSVSGTITVRSYDMMNEEQYYQFFLSVLDVYGFTVIPMDNNVLKIIR
E.coli_O157        EFSANFKDTDIQEFINTVSKNLHKTVIINPDVQGTITVRSYDMLNEEQYYQFFLSVLDVYGFAVVDMHNGILKVVR
K.pneumoniae       EFSASFKGTDIQEFINTVSKNLNKTVIIDPSVRGTITVRSYDMLNEEQYYQFFLSVLDVYGFAVINMNNGVLKVVR
V.fischeri         DFSASFKGTDIQEFINIVGRNLEKTIIVDPSVRGKVDVRSYDLLNEEQYYQFFLSVLEVYGYAVVEMDSGILKVIK
V.cholerae         EFSASFKGTDIQEFINIVGRNLEKTIIVDPSVRGKVDVRSFDTLNEEQYYSFFLSVLEVYGFAVVEMDNGVLKVIK
                      β1        α1        β2        β3        α2        β4        β5
```

Figure 5.20 HADDOCK generated OutC-HRF3 and OutD-N0 interaction module. (A) OutC-HRF3 OutD-N0 complex in cartoon representation. (B) individual protein interaction surface plot. The OutC-HRF3 (cyan) residues involved in interaction is shown in red. OutD-N0 (green) residues involved in the interaction are shown in orange. (C) GspC-HR and GspD-N0 sequence alignment.

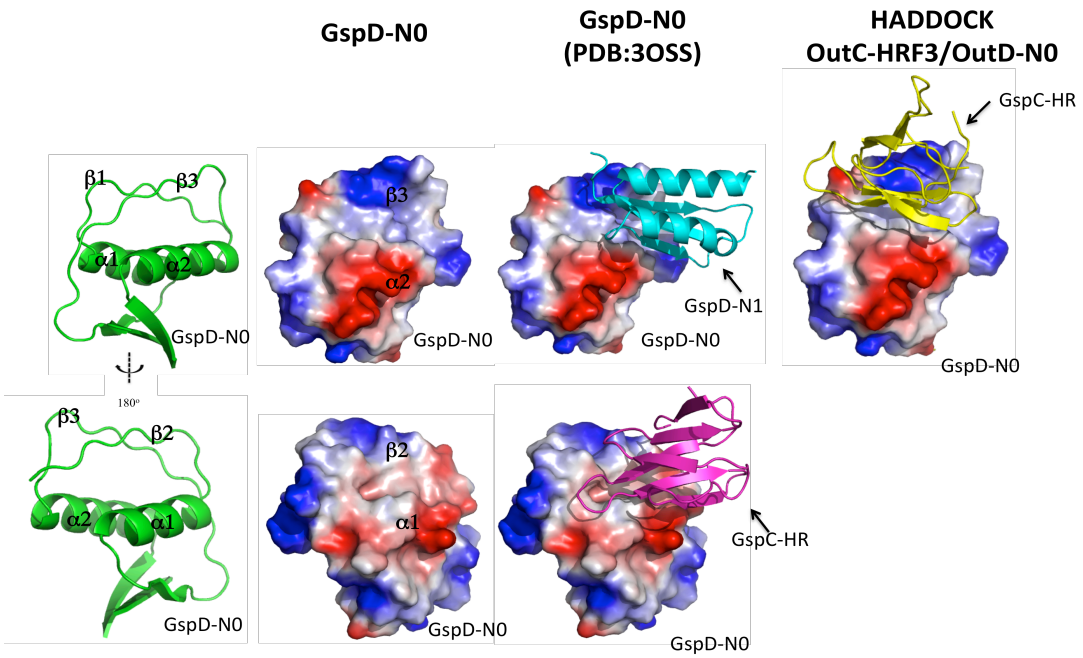


Figure 5.21 Comparision of GspC/GspD (PDB:3OSS) structrue and HADDOCK OutC-HRF3/OutD-N0 structure.

In recent paper, the available GspD crystal structure is fitted on to EM envelope (Reichow, Korotkov et al. 2010). However, there are a few limitations to this study. Firstly, the current available GspD periplasmic region structure is solved in the presence of nanobody. The presence of the nanobody is creates contact surface to form a crystal lattice. However, the nanobody contacts both N0 and N1 domain with a total buried surface of 1896Å^2. This contact is large compared to the N0 and N1 contact surface of 1180Å^2. The N0 N1 is treated as a complex in generating the dodecamer. The equivalent N0 N1 domain structure present in the secretin of type III secretion secretin showed a high sequence similarity but different orientation (Spreter, Yip et al. 2009). So N0 and N1 domains probably may not have the relative orientation seen in the nanobody crystal structure and further investigation is required. Secondly, the fitting ignored the domain integration between different monomers, as observed in Type IV secretin, there is massive domain integration present, the domain not only interacts the neighboring subunit but also the domains some distance away.

5.7 Interaction study of OutC-HRF3 and OutD-N0 by thermofluor

Thermofluor is a high-throughput protein stability test. In this case, protein OutC-HRF3 and OutD-N0 were used to test the stability of the proteins in different buffer conditions. The buffer that gave the higher melting point tends to provide the protein with better stability.

Protein folding/unfolding was monitored using the dye SYPRO-orange (Invitrogen). The fluorescence of SYPRO-orange is highly quenched in an aqueous environment. As the protein unfolds, hydrophobic surfaces that are buried in the native protein become exposed to solvent and SYPRO-orange binds to these hydrophobic sites and the fluorescence signal from the dye increases.

OutD-N0 alone gave good signals at the lower pH values of 4.5 and 5.5. Whereas the OutC-HRF3 did not show as good affinity for SYPRO-orange, and gave at least 3-4 fold less signal than OutD-N0 (Figure 5.20 A, B). OutD shows melting around 60^{O}C. OutC-HRF3 shows melting around 50^{O}C. This result is consistent with the melting experiment done by NMR.

Thermofluor experiments used buffers made for crystallization purposes. The highest melting temperature for OutD-N0 was measured at pH 9.5. At pH values above 7.5, the amide proton exchange rate is high and it is difficult to study interaction by NMR. The next highest melting temperature was pH 5.5. Hence, to investigate the protein-protein interaction, a buffer around pH 5.5 was chosen to optimize the wavelength for measured signal and measurement time. Titration at pH 5.5 with or without 150mM NaCl gave very similar results to titration at pH7.0.

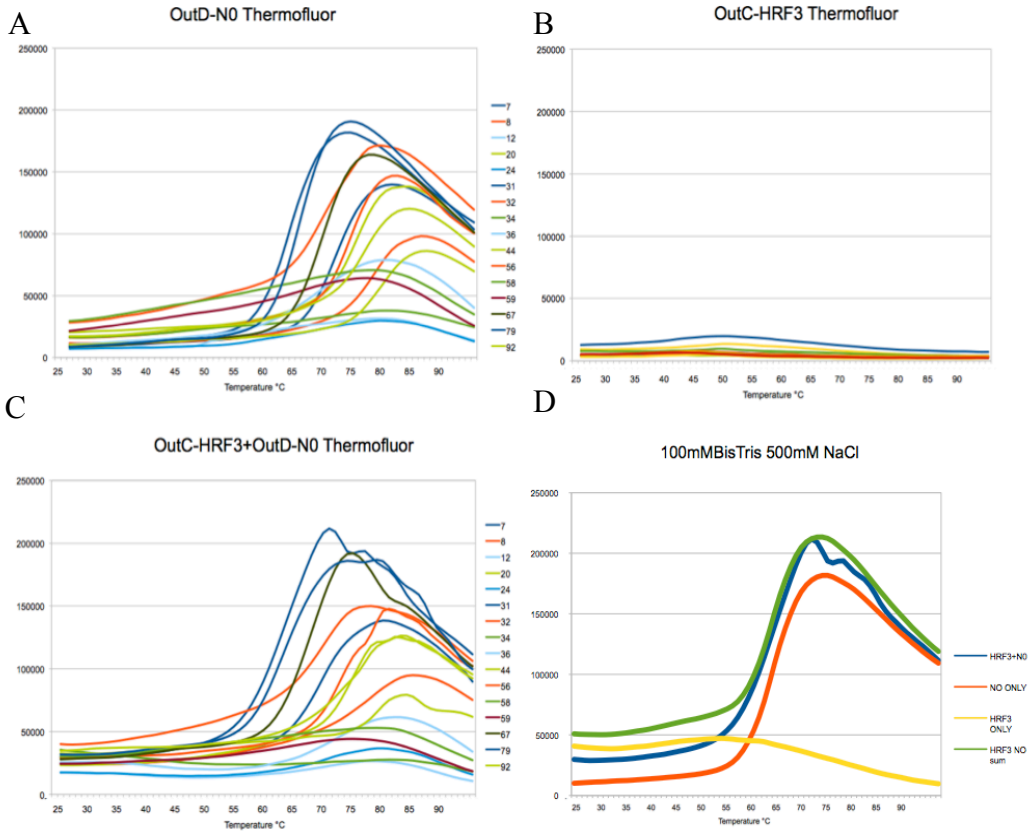


Figure 5.22 Thermofluor plots. (A) Thermofluor plot of OutD-N0. (B) Thermofluor plot of OutC-HRF3. (C) Thermofluor plot of a mixture opf OutC-HRF3 and OuD-N0. (D) Thermofluor plot of all the proteins using condition C2 (100mM BisTris 500mM NaCl pH5.5).

When OutC-HRF3 and OutD-N0 were mixed the traces develop a shoulder and the melting temperature is shifted 10°C higher. These provides further evidence for the two proteins are interacting.

5.8 OutC-HRF3 and OutD-N0 interaction by His Trap-pull down assay

The NMR titration results suggest the OutC-HRF3 and OutD-N0 interaction is very weak. Dr Shevichik (Lyon) provided evidence that when OutC and OutD expressed together, OutD-N0 can be pulled out with several GST-OutC constructs ($OutC^{144-158}$, $OutC^{128-172}$) using a glutathione column. In an attempt to stimulate the OutC and OutD complex formation, OutC-HRF3 and OutD-N0 co-expression construct was made (by

Dr Shevichik), which co-express N-terminus GST tagged OutC-HRF3 and C-terminus His-tagged OutD-N0.

Unfortunately GST tagged OutC-HRF3 and the majority of OutD-N0 produced are insoluble. Hence, urea was added to solublilize both proteins. Buffer was then gradually added into urea suspension stimulate protein refolding, this might mimic the *in vivo* chaperone activity. Both OutC-HRF3 and OutD-N0 can be solubilised and refolded into soluble protein in buffer without detergent. Since urea is present in the solution and high salt is used in the washing solution to eliminate the contamination and non-specific bindings, a Ni column was selected for the pull down assay. As shown in Figure 5.22(A) OutC-HRF3 can be pulled out with His tagged OutD-N0 using this method. This indicates that OutC-HRF3 and OutD-N0 refolding was successful and both protein can form a complex through refolding.

It is noticed that the eluate from the Ni-column contains more His tagged OutD-N0 than pulled out OutC-HRF3 partner and OutC-HRF3 is also present in quantity in the Ni column flow-through. This suggests the interaction is weak, which is further supported by the observation that the complex dissociates during size exclusion chromatography (Figure 5.21B). The pull-down assay however gives further independent support for the OutC-HRF3 interaction with OutD-N0.

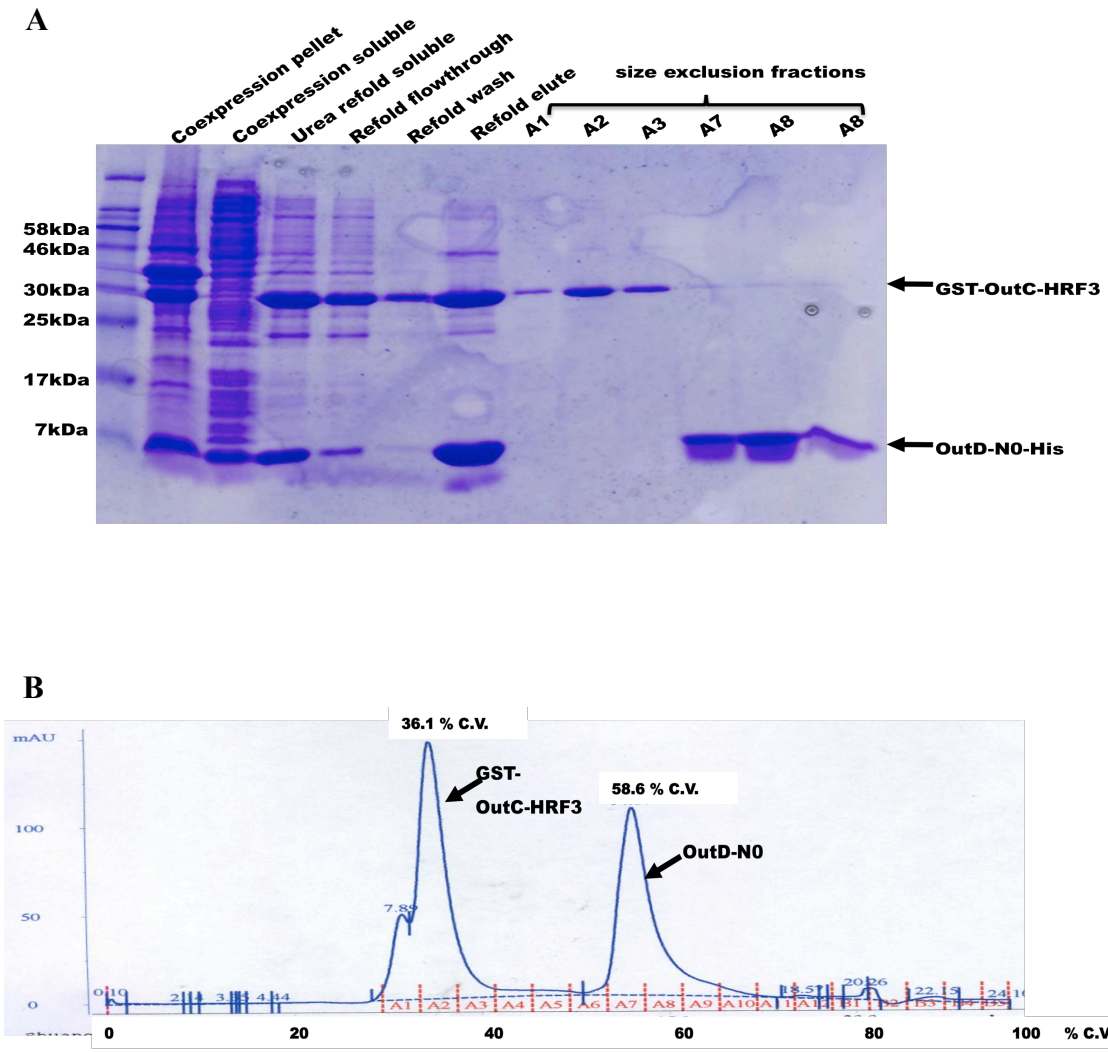


Figure 5.22 OutC-HRF3 OutD-N0 Ni column pull down assay. (A) SDS-PAGE analysis. (B) Size exclusion trace of Ni column eluate, using a Superdex 75 column.

5.9 Summary and discussion

In this Chapter, several techniques were employed to explore the interaction between OutC-HRF3 and OutD-N0. The experiments are all first carried out at 20mM Tris pH7.0 and later repeated at 50mM NaAc 150mM NaCl pH5.5. These conditions are all biologial relevant therefore the results reflect the interaction *in vivo*.

One of the reason for study the GspC/GspD interaction by NMR is that the interaction is very weak, with estimated K_d in the mM range. Since *in vivo*, OutD is a dodecameric complex, the weak interaction observed with OutC-HRF3 and Out-N0 subunits is anticipated to be significantly stronger when they interact cooperatively in the oligomer. The affinity between the monomers may not reflect the actual affinity between OutC and OutD *in vivo* in the context of the assembled secretion system. The studies described here give an indication of which residues could be involved in the interaction and allow a model of the interaction to be proposed. Since obtaining the HADDOCK model, our collaborator Dr Shevchik et al. have successfully isolated OutC-HRF3^{G54C}/OutD-N0^{S39C} complex *in vivo* through disulphide cross-linking. This validates the proposed model of interaction.

The recent GspC/GspD structure (Korotkov, Johnson et al. 2011) structure suggest that GspC interacts with a different hydrophobic patch on GspD-N0. The difference between models are mainly caused by the presence of GspD-N1 pre-occupied the GspC-HR site we proposed. When the site is pre-occupied with GspC-HR, first the GspD-N1 do not seems to compete with GspC-HR (from this work, Figure 5.17). This suggest the affinity between GspD-N0/GspC-HRF3 and GspD-N0/GspD-N1 are comparable. It is also shown the force between these interactions are through hydrophobic surfaces.

These properties make these interactions are inter-interchangeable through rearangement between these subunits during the conformational changes.

Chapter 6 Interaction between OutS and OutD

6.1 Overview

The other protein known to interact with the secretin (GspD) of the type II secretion system is the pilotin (Gsps), whose role is to guide the secretin to the outer membrane. Since OutC-HRF3 shares a closely similar structure with the type IV pilotin PilP, the T2SS pilotin is of additional interest.

I initiated the GspS structure determination and involved in early stage of structure determination, including protein purification, DLS, CD and crystal screen. The project was then passed to Saima Rehman. The crystal structure of the pilotin was solved in our Laboratory in December 2010, I was also involved in the X-ray data collection at ESRF, Grenoble. Our collaborator Dr.Shevchik provided constructs to overproduce the pilotin and the C-terminal part of the secretin from which it was known to interact. The plasmids harbored the *Erwinia chrysanthemi* OutS (pilotin) and the C-terminal 62 residues of the (OutD-Cter secretin). The OutD C-terminal 62 residues had previously been shown to involve in the interaction with OutS (Shevchik, Robert-Baudouy et al. 1997). In this Chapter, the interaction between pilotin OutS and secretin OutD-Cter is studied by CD and NMR spectroscopy.

6.2 Bioinformatic studies of OutS and OutD-Cter

GspS (called OutS and PulS in *Erwinia* and *Klebsiella* respectively) is the only identified pilotin in the type II secretion system. It is a liproprotein transported by Lol system from the inner membrane to the outer membrane. GspS interacts with secretin, GspD to achieve correct targeting and insertion of the secretin into the outer membrane.

In the absence of pilotin PulS, secretin PulD is mislocated to the inner membrane (Guilvout, Chami et al. 2006). The only available type III secretion pilotin revealed a predominant β strand structure forming the cracked β barrel structure (Lario, Pfuetzner et al. 2005). Lario and colleagues suggested that all pilotins would have the same fold or maintain critical structural fratures (Lario, Pfuetzner et al. 2005).

The results of previous sequence comparisons using GspS and its interacting partner OutD-Cter have been published (Shevchik, Robert-Baudouy et al. 1997). Here,we can see from secondary structure prediction using J-pred (http://www.compbio.dundee.ac.uk/www-jpred/) that the GspS structure is predicted to be mainly helical (Figure 6.2A). Since the only known pilotin of similar size (~10kDa, i.e.PilP, MxiM) are all have β-strand dominated structure (Lario, Pfuetzner et al. 2005; Golovanov, Balasingham et al. 2006), the GspS structure is a curiosity and could be a novel pilotin structure. Therefore, the crystal structure of OutS was determined in our lab. The OutS structure revealed a novel nested α-helical architecture with one helical hairpin perpendicular to the second providing a concave surface for secretin-binding. The most similar arrangement of OutS α-helices is found in P40 nucleoprotein with a DALI score of 6.0 and sequence identity 5% (Figure 6.2). OutS has a concave surface that plausibly could bind secretin peptide. When OutS is compared with p40 nucleoprotein, the corresponding surface in P40 nucleoprotein is occupied by a helix. This suggests that OutS could bind to a helical structure present in the OutD-Cter peptide. The secondary structure of OutD-Cter is predicted to be three helices which agrees well with this hypothesis (Figure 6.2).

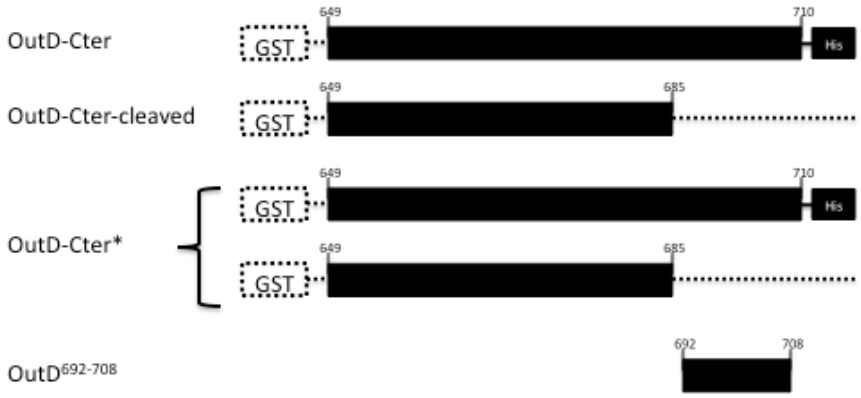


Figure 6.1 Schematic diagrams of OutD C-terminal derivatives used in this Chapter. The dotted line represents region is expressed in cell but later cleaved-off by the protease.

A

```
Secondary -------HHHHHHHHHHHHHHH-----  --------  -------HHHHHHHHHHHHHHHHHHHH------
                        10        20        30        40        50        60
                         |         |         |         |         |         |
OutS_Echr ---MHVSSLKVVLFGVCCLSLAACQTP--APVKNTAS--RSAASVPANEQISQLASLVAASKYLRVQCERSDL
OutS_Pcar ---MSLSLSKFSALTLLCVTLAGCQQTGTSAHKGTVAGQSSAVTVSPNDQLNQLSSLVAATRYLKSKCNRSDL
GspS_Ecol MMGNILKKLNCIASLLVLVTISGCHQS--PSIHKQAT-------VPPSEQLEQMASIVSATRYLKMRCNRSDL
PulS_Koxy ------MRNFILFPMMAVVLLSGCQQN--RPTTLSPA-------VSGQAQLEQLASVAAGARYLKNKCNRSDL
GspS_Ymol ----MLSTTRKLFWFLPLVLLTGCQQP-MNKTIKPTA-------Q---QQIKQLSALVAGAHYLQKNCQRAEV
GspS_Spro ---MQRMLRSFSFAAAAVVLLAGCQQTALQPQTQPSL-------T---AQLDQLSALLAGSQFLRQHCARTDI

Secondary  --HHHHHHHHHHHHH----H  HH----HHHHHHHHHHHH----HHHHHHHHHH-HHHHHHHH---
             70        80         90        100       110       120       130
              |         |          |         |         |         |         |
OutS_Echr  PDDGTILKTAVNVAVQKGWDT--GRYQSLPQLSENLYQGLLKDGTPKATQCSSFNRTMTPFLDAMRTVR--
OutS_Pcar  PDDATVMNVALTVAKQRGWNV--ASYQALPQRSESLYQGLLKDSTPKETQCSEFNRTLTPFIDAIRSRG--
GspS_Ecol  PDEQSILNVANRIAIGKGWQS--LTQEDIRKHSDDIYVRLTRDSTPEYIKCREFNRRLVPFIGELLARGRG
PulS_Koxy  PADEAINRAAINVGKKRGWAN--IDANLLSQRSAQLYQQLQQDSTPEATKCSQFNRQLAPFIDSLRDNK--
GspS_Ymol  PDEAVLLKTARSLAASRHWDTRAPAYKLLGEQSQARYQALVKENETDKSMCTELNLLMVDFVDEAQRNIK
GspS_Spro  PDDASLQRSAIGMAQQRGWNTQPAEYRQLPVRAQQRYQQLQQDGTPLQQKCAALNTSTARFIAAAQSDARQ
```

B

```
Secondary  ---------HHHHHHHH---HHHHHHH------   ---------- -----HHHHHHHHHHHH------
             650       660       670          680       690       700       710
              |         |         |            |         |         |         |
OutD_Echr  TIIRDPGQFQEASINKYRSFNNEQQQQRGEGNG---VLDNNTLRLS-GGNTYTFRQVQSSISDFYKPEGR
OutD_Pcar  SIIRDRSQYQSASASKYHSFNAEEEKQREANGGKANLLDNDLLRLPEGGNAYTFRQVQSSIVAFYPAGGK
GspD_Ecol  TIIRERDGFRHASAEKYQSFNQEQVQSRGKETT-ALTLNEEQLRLSPDQDDTAFRKVKAAIAAFYAQEM-
PulD_Koxy  TVIRDRDEYRQASSGQYTAFNDAQSKQRGKENND-AMLNQDLLEIYPRQDTAAFRQVSAAIDAFNLGGNL
GspD_Ymol  TIIREQDSYTDLSEQRLDKFQQEQNRDRPASGQ---RINENLNNIL--SDGRSLQDLRNDVSAFYSKGA-
GspD_Spro  TVIRSTEEYAQESMRKAARFRSLDAQPLPLKTH----AEAQLDRVG-STENGTFQRIQRQIDAFYPRGEL
```

Figure 6.2 Family sequence alignment of GspS (A) and the C-terminal region of cognate GspD (B). Position of the α-helices is indicated by pink H (predicted by Jpred for GspD). Shown are the GspS and GspD homologs of *Erwinia chrysanthemi* 3937, Q01567 and Q01565; *Pectobacterium carotovorum*, C6DAR0 and C6DAQ5; *Escherichia coli* O157:H7, Q7BSV3 and Q9ZGU0; *Klebsiella oxytoca*, P20440 and P15644; *Yersinia mollaretii*, C4S9G3 and C4S9F5; *Serratia odorifera*, D4E1I4 and A8GJQ5. Identical residues are in red, strongly similar and weakly similar residues are in green and blue respectively.

6.3 Expression and purification of OutS

To produce sufficient quantities of soluble GspS for structure study, a recombinant form of *Ewinia chrysanthemi* OutS, which corresponding to residue 27 to 133 expressed in pET20b vector was kindly provided by Dr Vladimir Shevchik (University of Lyon). Residues 1 to 26 corresponding to a N-terminal lipoprotein sequence were eliminated in this construct. Instead, the pET20b vector provides a PelB sequence to the recombinant protein, which facilitates the protein's transportation to the periplasmic region where the PelB sequence is cleaved during the transport process. It is crucial for recombinantly produced OutS to be transported to the periplasmic region, which mimics its native functional environment and provides a non-reducing environment for OutS internal disulphide bond formation and protein folding. The transportation of OutS to periplasmic region also made OutS purification straight forward by osmotic shock (described in 2.2.3) (Fries, Ihrig et al. 2007))

OutS pET 20 vector was transformed into *E.coli* BL21 (DE3) cells. The expression can be induced by 500μM of IPTG. OutS production in LB media can reach about 5mg per liter. However, OutS cannot be produced in M9 media following similar procedure. This is owing to the transportation of OutS from cytoplasm to outer membrane is operated by ATP hydrolysis which is required for Sec system function (Driessen 1992; Pugsley 1993). Limited nutrient in the M9 media could not support a functional sec system. Several altered methods were tried and OutS production can be achieved in M9 media with the addition of 20mg ATP per liter during induction.

OutS can be extracted by single step sucrose osmotic shock (described in 2.2.3). The separated sucrose soluble protein was then concentrated and purified on size exclusion

column. OutS elutes at 13.38 ml on S75 column, consistent with OutS being a monomer in solution (Figure 6.3).

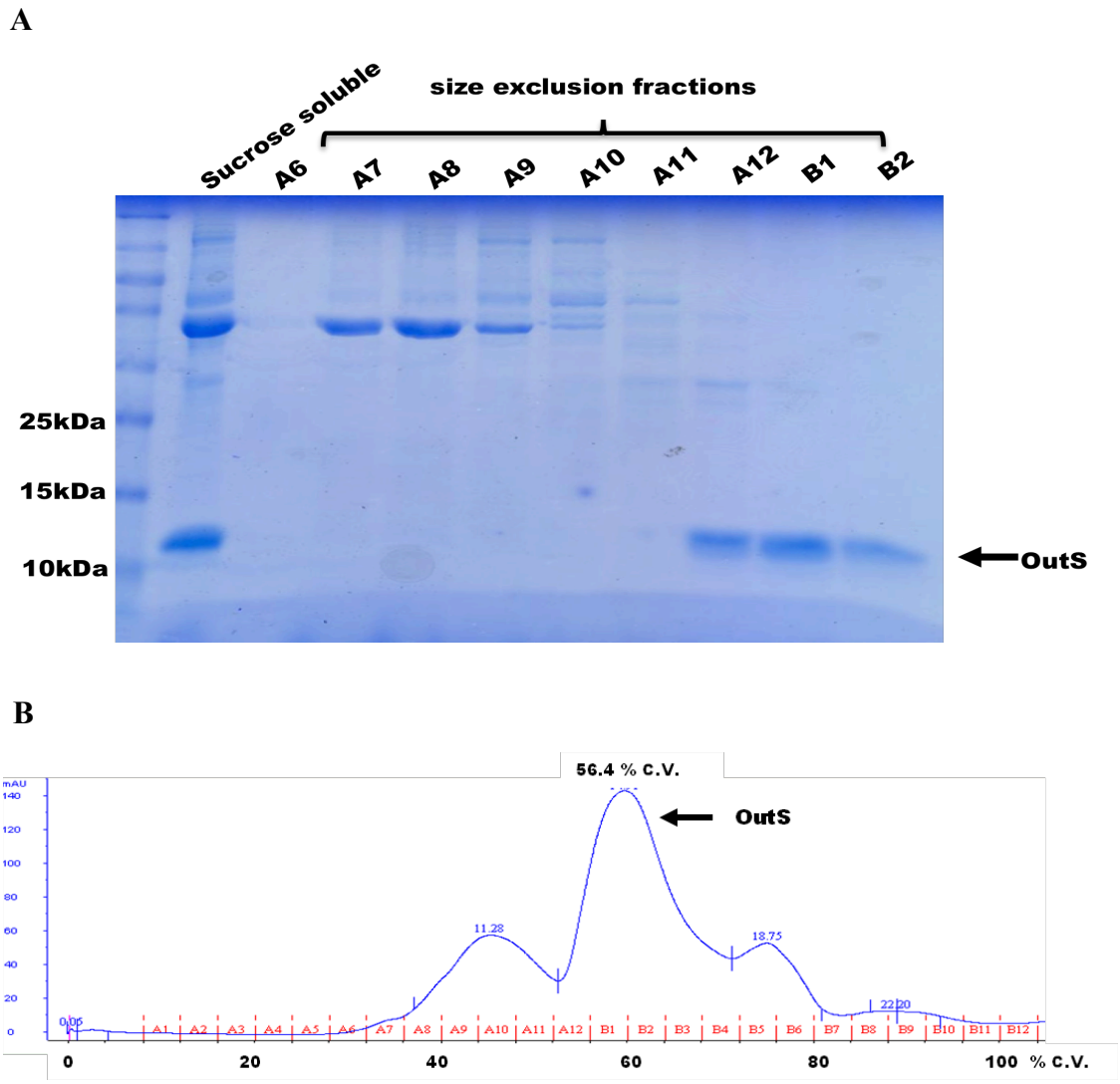


Figure 6.3 Purification of OutS (A) Purification of OutS (B) OutS S75 column size exclusion trace at 280nm. OutS elution peak as indicated.

6.4 Expression and purification of OutD-Cter

For structural studies OutD-Cter, a recombinant form of the OutD C-terminal 62 residues expressed in pGEX-6P-3 vector was kindly provided by Dr Vladimir Shevchik (University of Lyon). (Shevchik, Robert-Baudouy et al. 1997). The pGEX-6p-3 vector provides an N-terminal GST tag for the recombinant protein, which helps with protein

solubility as well as aiding purification. OutD-Cter also contains an N-terminal non-cleavable His tag for further purification.

6.4.1 Purification of OutD-Cter with GST tag*

OutD-Cter* was purified using a glutathione column followed by a size exclusion column (described in 2.2.3 and 2.2.5) allowing OutD-Cter* can be produced in high purity.

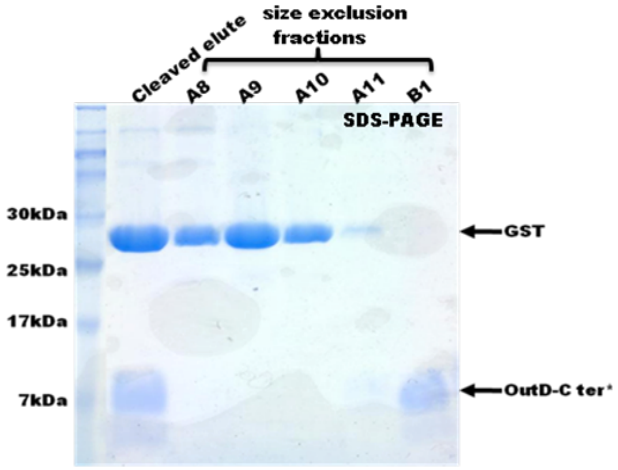
The OutD-Cter* band is smeary on SDS-PAGE (Figure 6.4A). It turns out OutD-Cter* produced by this method is a mixture of OutD-Cter and OutD-Cter-cleaved (Figure 6.4C). However, the OutD-Cter* produced is found to be useful for NMR studies.

6.4.2 Purification of OutD-Cter with His tag and GST Tag

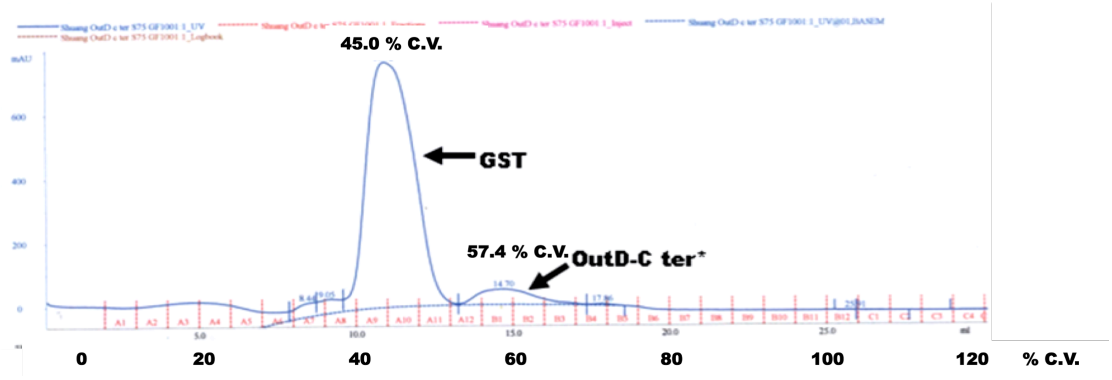
Ideally, a pure preparation of OutD-Cter is required for circular dichroism and NMR studies. Since OutD-Cter-cleaved is cleaved from the C-terminus and is missing the non-cleavable C-terminus His tag further purification can be achieved using the His tag.

As shown in Figure 6.5A OutD-Cter full can be purified to a relatively high purity after Ni affinity column, glutathione column and size exclusion column purification. The purified protein is still prone to proteolysis and likely to be cleaved at R685 site and produce OutD-Cter*. Therefore, the purified protein was stored at -80^{O}C to minimize any protease activity and protein degradation.

A



B

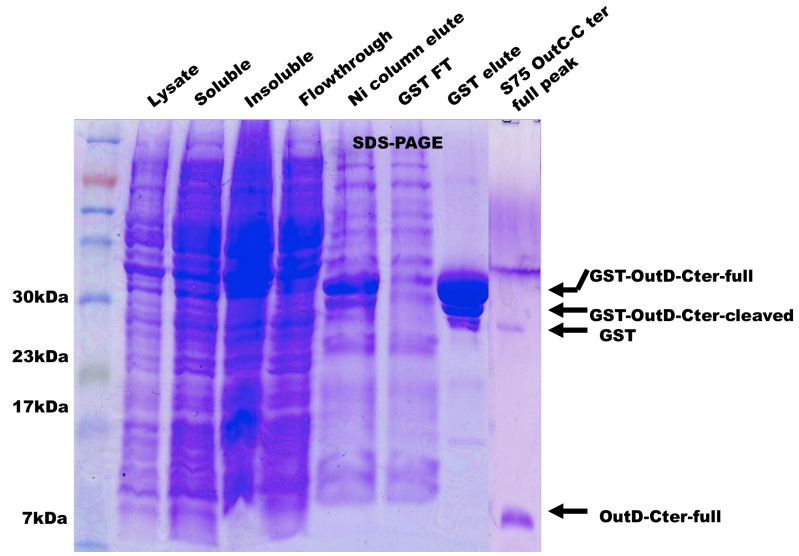


C

```
                1         10649                                685                    710
                |         | |                                  |                      |
OutD-Cter       GPLGSPNSRVDPGQFQEASINKYRSFNNEQQQQRGEGNGVLDNNTLRLSGGNTYTFRQVQSSISDFYKPEGR
                 Tag
OutD-Cter-cleaved GPLGSPNSRVDPGQFQEASINKYRSFNNEQQQQRGEGNGVLDNNTLR
```

Figure 6.4 Purification of OutD-Cter* (A) SDS-PAGE analysis. (B) uv absorbtion of eluant from the S75 column. OutD-Cter* elution peak is indicated (C) Sequence of the OutD-Cter and its cleaved derivative OutD-Cter-cleaved sequence.

A



B

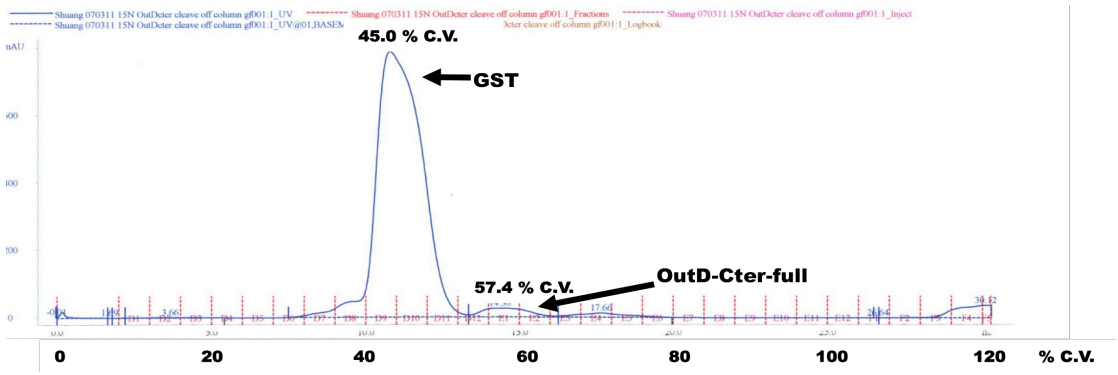


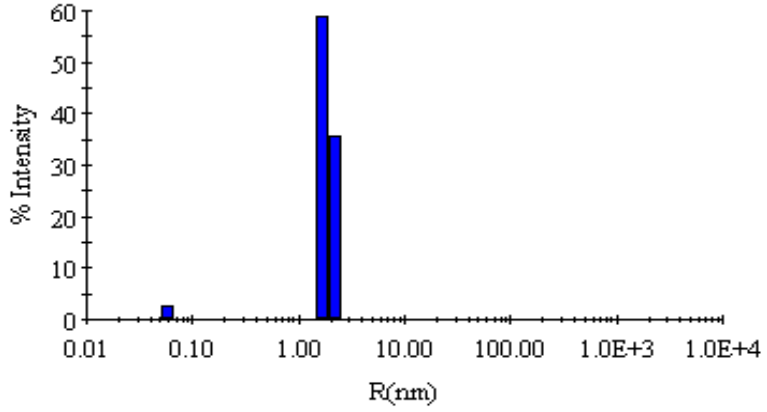
Figure 6.5 Purification of OutD-Cter (A) SDA-PAGE. (B) OutD-Cter S75 column size exclusion trace at 280nm. OutD-Cter elution peaks are indicated.

6.5 Biophysical studies of OutS and OutD-Cter

Dynamic Light Scattering (DLS) was used assess if the OutS produced was suitable for structural studies. Since there was no structural information about OutS and OutD-Cter, CD experiments were used to determine the secondary structure contents of both proteins individually and in complex.

6.5.1 OutS DLS

DLS was used to check whether the OutS produced is properly folded, not aggregating and suitable for crystallography studies. The result showed it behaved as a 15kDa protein with polydispersity of 13.9% this indicates that OutS produced by this method is folded and behaves as monomer in solution (Figure 6.6). It is generally accepted that proteins with polydispersity of less than 20%, have a very good chance of crystallizing. With 13.9% polydispersity, OutS is crystallized good candidate. Crystallization trails led to the growth of diffracting crystals from which the crystal structure was determined.



	Radius(nm)	%Pd	MW-R(kDa)	%Int
Peak1	0.05844	0	0	2.6
Peak2	1.88112	13.9	15	97.4

Figure 6.6 OutS DLS summary. OutS was measured at 10mg/ml in 20mM Tris pH7.0 at 20^{O}C.

6.5.2 OutS CD

The far UV-CD spectrum gives good indication of secondary structure. The OutS CD spectra (Figure 6.7) peaked at 190nm and dipped at 210 and 225nm, showing OutS is predominantly an α-helical protein. This CD result is consistent with the secondary structure prediction.

6.5.3 OutD-Cter CD

CD was measured using OutD-Cter to give a general idea of OutD-Cter secondary structure. In 20mTris, 150mM NaCl, pH7.0, OutD–Cter gives a typical disordered spectra with a dip at around 197 nm (Figure 6.8). To check whether OutD-Cter is capable of forming helical structure as predicted by J-pred, 2,2,2-trifluoroethanol (TFE) was added to OutD-Cter solution. TFE is an alcohol with low-dielectric constant and can induce stable conformation that resembles of the interior of a protein. In the presence of a low concentration of TFE (16%), part of OutD-Cter started to fold into α-helical structure (Figure 6.8). This provides support for the J-pred prediction.

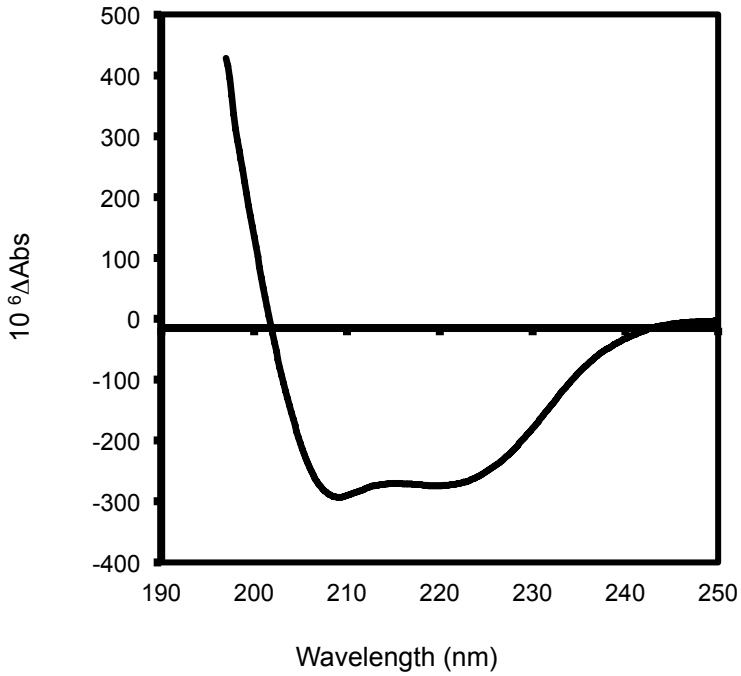


Figure 6.7 Far-UV CD spectrum of OutS . The spectrum was measured at 20^{O}C in 20mM Tris, 150mM NaCl pH7.0 using 2mg/ml OutS.

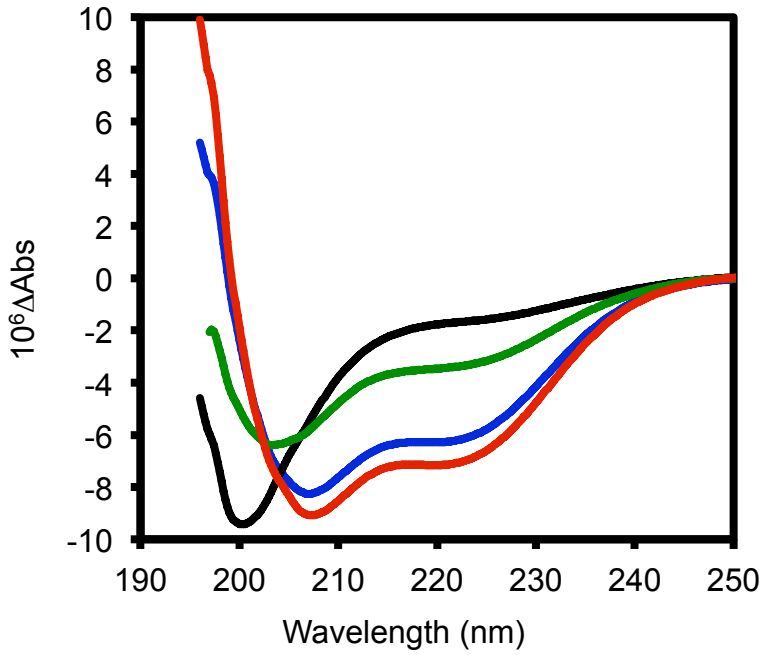


Figure 6.8 Far-UV CD spectra of OutD-Cter (black) and in the presence of 16% (green), 28% (blue) and 37% (red) of TFE. 0.05mg/ml (~1.1μM) OutD-Cter was used for the experiment and measured at 20°C in 20mM Tris 150mM NaCl pH7.0.

6.5.4 Interaction studies of OutS and OutD-Cter by CD

OutD-Cter is unstructured on its own and predicted to adopt a principally helical structure by J-pred. In the presence of TFE, OutD-Cter is able to form a helical structure. To determine if helical structure can be induced in OutD-Cter by the presence of OutS, CD spectra were recorded for the mixture of OutD-Cter and OutS. The CD spectra were measured using 0.044mg/ml (0.55μM) of OutD-Cter and 0.06mg/ml (0.55μM) of OutS in 20mM Tris, 150mM NaCl, pH7.0, separately. As suggested by NMR titration, the OutS and OutD form complex in a 1:1 ratio (see section 6.7), an equal volume of OutD-Cter and OutS sample was taken and mixed to record the CD spectra. The OutD-Cter and OutS mixed sample with the final concentration of both proteins halved, therefore data were mulitiplied by two to compensate the signal loss by dilution. In Figure 6.9, OutD-Cter and OutS spectra are shown; the red spectrum is the

additive sum of the two spectra. This shows the OutD-Cter and OutS mixture spectra in the absence of the secondary structure changes. The actual spectrum of the OutD-Cter and OutS mixture is shown in blue. The actual mixture spectrum shows a further dip at 210 and 225nm and peaked higher at 190nm. To quantify the secondary structure of the OutS and OutD-Cter mixture spectrum, the actual measured spectra has 41.5% (77 residues) of helical and 7.2% of β-strand content, whereas the sum of the independent OutS and OutD-Cter spectra has 34.8% (65 residues) of helical content and 7.7% beta strand content. The quantified result provided strong evidence of new helical formation, when mixing the two proteins together. As the OutS structure is solved, the residues in helical conformation can be determined as about 68 out of total 107 residues in OutS construct. In the mixture of OutS (107 residues) and OutD-Cter (87 residues) (1:1), assuming the secondary structure of OutS does not change but that of the unstructured peptide does, the OutS secondary content should be around 36% (68 out of 185 residues) and is consistent with 34.8% measured. As the K_d of the interaction was not measured, the percentage of complex formation was unknown. Assuming neglidgible change in the OutS secondary structure content, at least 12 residues (assuming 100% complex formation) from OutD-Cter form helical structure when the peptide is bound to OutS.

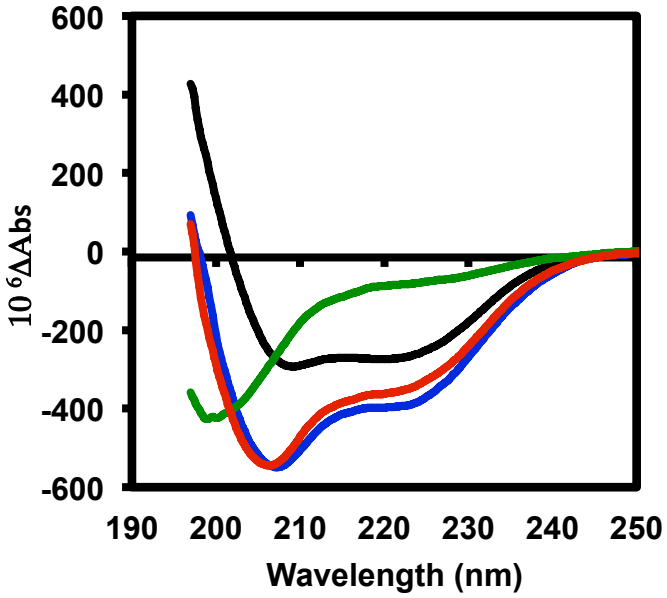


Figure 6.9 Far-UV CD spectra of OutS and OutD-Cter mixture. Far UV spectrum of OutS (black), OutD-Cter (green) and OutS and OutD-Cter mixture (blue). The additive sum of OutS and OutD-Cter is shown in red. OutS and OutD-Cter both measured at 0.55μM. The mixture of OutS and OutD-Cter was made by mixing equal amount of OutD-Cter and OutS samples and measured at same condition. As the OutS and OutD mixture both diluted by half, the CD data were multiplied by two compensate the dilution factor.

6.6 NMR study of OutS and OutD-Cter

CD data gives preliminary evidence of the OutS and OutD-Cter interaction. NMR spectroscopy was used to gain more detailed information. Initially OutS and OutD-Cter were measured individually to pave the way for the interaction study.

6.6.1 OutS 2D HSQC spectrum

The OutS spectrum is well dispersed and the peaks are sharp (Figure 6.10). This shows that OutS behaves as homogenous monomer. The number of peaks shown on the spectrum reflects the actual backbone amide protons of OutS. Therefore, the experimental condition is useful for further experiment, such as HN based 3D experiments for backbone assignments.

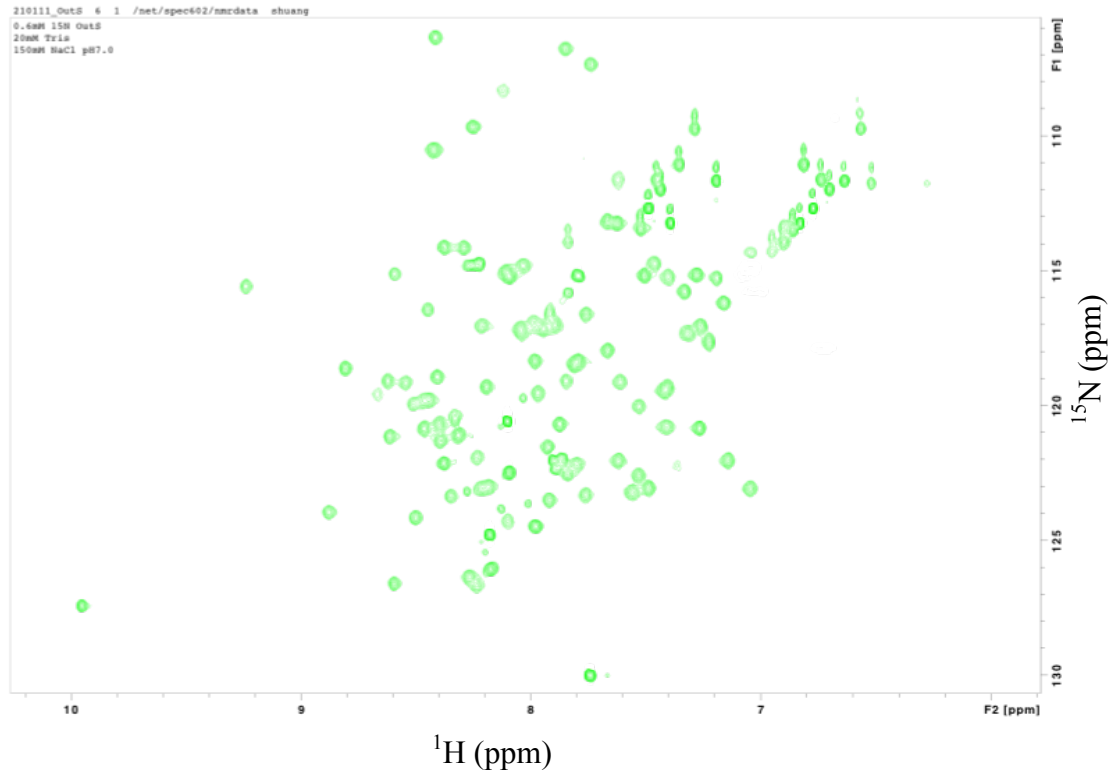


Figure 6.10 ^{1}H-^{15}N HSQC spectra of OutS. The spectrum was acquired at Bruker 600MHz at 15^{o}C in 20mM Tris 150mM NaCl (pH 7.0) and 10% $^{2}H_2O$.

6.6.2 OutD-Cter 2D HSQC spectrum

According to OutD-Cter* ^{1}H-^{15}N HSQC spectra, all backbone amide protons appear in the region between 7.5ppm to 8.5ppm. This indicates that the OutD-Cter peptide is unstructured in solution. The number of backbone amide protons appearing in the 2D HSQC is correct for the number of backbone amide proton present in the protein.

The unstructured nature of the OutD-Cter peptide makes it susceptible to protease activities *in vivo*. With a closer look at the OutD-Cter ^{1}H-^{15}N HSQC, two sets of peaks can be seen to be present in the spectrum, one set of very intense peaks and one set of weaker peaks. During the backbone assignment, peaks with strong intensities are residues upto residue R685, which indicates there is a cleaved derivitives present in the OutD-Cter*. It is no surprise as arginine is the cleavage site for many proteases. From the intensities of the peaks after R685, the full length OutD-Cter protein present in the

sample is much less than the cleaved version. It is also observed that the cleavage is carried out *in vitro*. It has been observed that the full length OutD-Cter is completely cleaved into OutD-Cter cleaved after 48hours at 15^{o}C (Figure 6.11 A). The cleaved off 26 residue peptide are relatively insoluble as precipitation is observed in the sample and no additional peaks are observed for the peptide in the HSQC spectrum.

6.7 Interaction between OutS and OutD-Cter*

The OutS and OutD-Cter interaction was initially shown by Dr. Shevchik (Shevchik, Robert-Baudouy et al. 1997) using a pull down assay. NMR titrations were used to better understand the interaction at a molecular level.

Cross titrations between OutS and OutD-Cter* were carried out. The labeled protein was titrated with unlabeled partner and the structural changes of the labeled protein during titration were monitored by recording the 2D HSQC spectrum. When OutD-Cter was titrated into ^{15}N labeled OutS, about a third of OutS peaks showed shifting (Figure 6.12). On the other hand, when titrating OutS into ^{15}N OutD-Cter*, only a few of the OutD-Cter* peaks shifted, while the majority of OutD-Cter* remained unchanged (Figure 6.13). From the titration result, it is likely that OutS interacts with a small patch on the OutD-Cter peptide, but the binding of the peptide has effects throughout the OutS

.

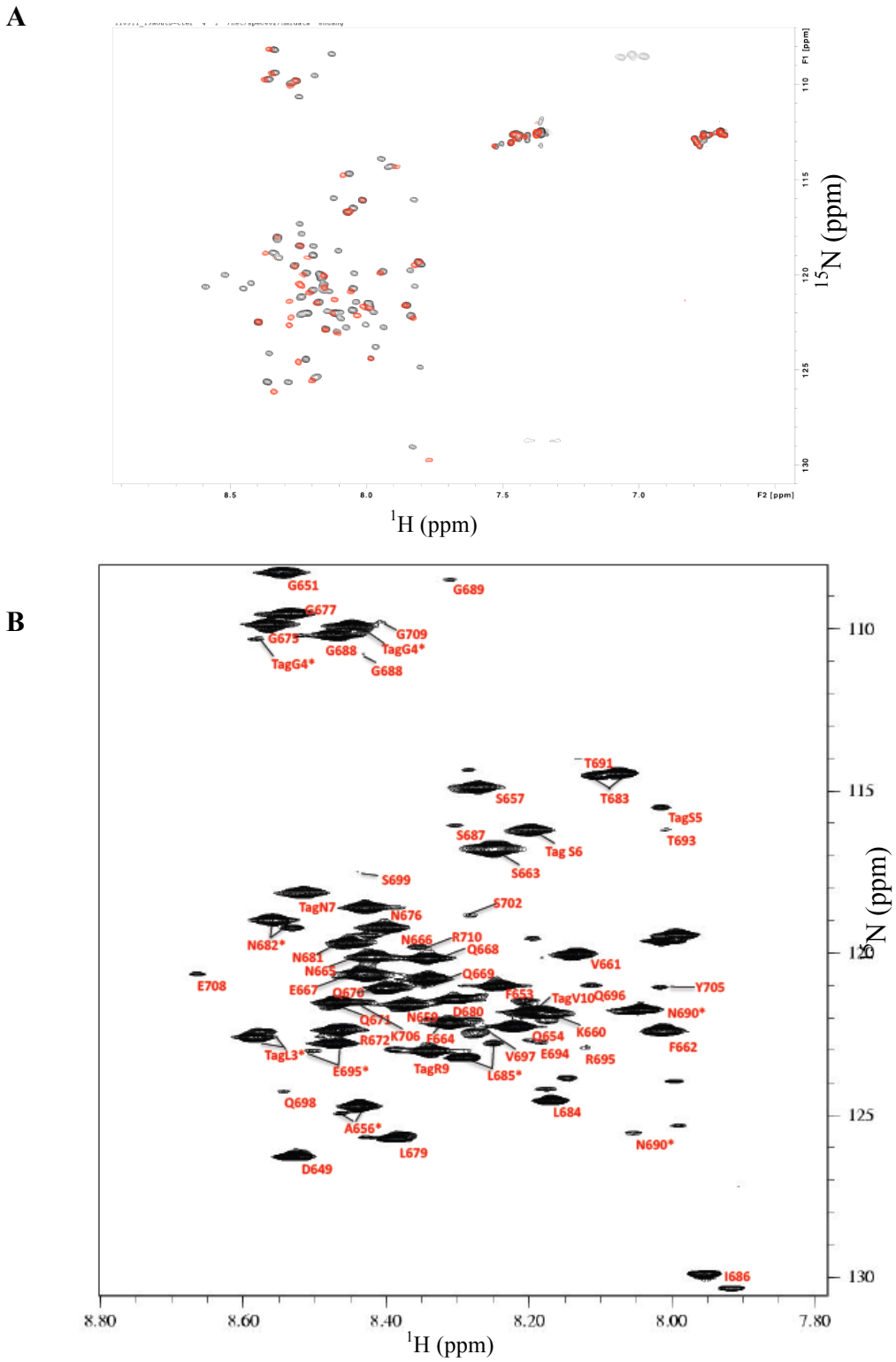


Figure 6.11 ^{1}H-^{15}N HSQC spectra of OutD-Cter* and backbone assignment (A)^{1}H-^{15}N HSQC spectra of OutD-Cter(Black) and OutD-Cter cleaved (Red). (B) OutD-Cter* backbone amide proton assignment. The OutD-Cter-cleaved sample is obtained by leaving OutD-Cter sample at 15^{O}C for 48 hrs. Spectra was acquired at Bruker 700MHz at 15^{o}C in 20mM Tris 150mM NaCl (pH 7.0) and 10% $^{2}H_2O$.

The interaction between OutS and OutD-Cter appears to be strong. Upon titration of the unlabeled partner protein, a new and well dispersed set of signals were observed in the ^{15}N-HSQC spectrum of the OutS (Figure 6.12). The peaks corresponding to the "apo" form became weaker and a new set of spectrum corresponding to the "bound" form appeared and become more intensified. The co-presence of the "apo" and "bound" form showed that the interaction between OutS and OutD-Cter behaves as slow exchange and indicates the tight binding between the two proteins. Since the titration experiments used 25μM labeled OutS with addition of 1.3 equivalents of OutD-Cter, all the OutS shifted to the "bound" state, the stoichiometry of the complex would be a 1:1 ratio.

The ^{1}H-^{15}N HSQC spectra of OutD-Cter*, a mixture of OutD-Cter and OutD-Cter-cleaved, showed poorly dispersed amide proton spectra, diagnostic of a predominantly unstructured peptide (Figure 6.11). Since OutD-Cter was produced recombinantly, it is co-produced with an *in vivo* proteolysis fragment OutD-Cter-cleaved. As OutD-Cter-cleaved was present predominantly in the OutD-Cter sample, the residues represent this region showed intense peaks (Figure 6.11 and Figure 6.13). When titrating, unlabeled OutS in to OutD-Cter*, the intense peaks, which are residues correspond to OutD-Cter-cleaved, all remained unchanged, (Figure 6.13A). Whereas the residues behave in a slow exchange regime are all the peaks with relatively weak intensities, which are residues on OutD$^{686-710}$ region (Figure 6.13). exchange in the titration. Hence, the interaction region on OutD-Cter is further located on OutD$^{691-710}$. As many of residues located on OutD$^{691-710}$ present in the crowded region of the ^{1}H-^{15}N HSQC spectra (Figure 6.13A) and are over shadowed with intense peaks from OutD$^{649-685}$ region the E708 is the last observable shifted peak. To confirm that the OutD-Cter-cleaved, does

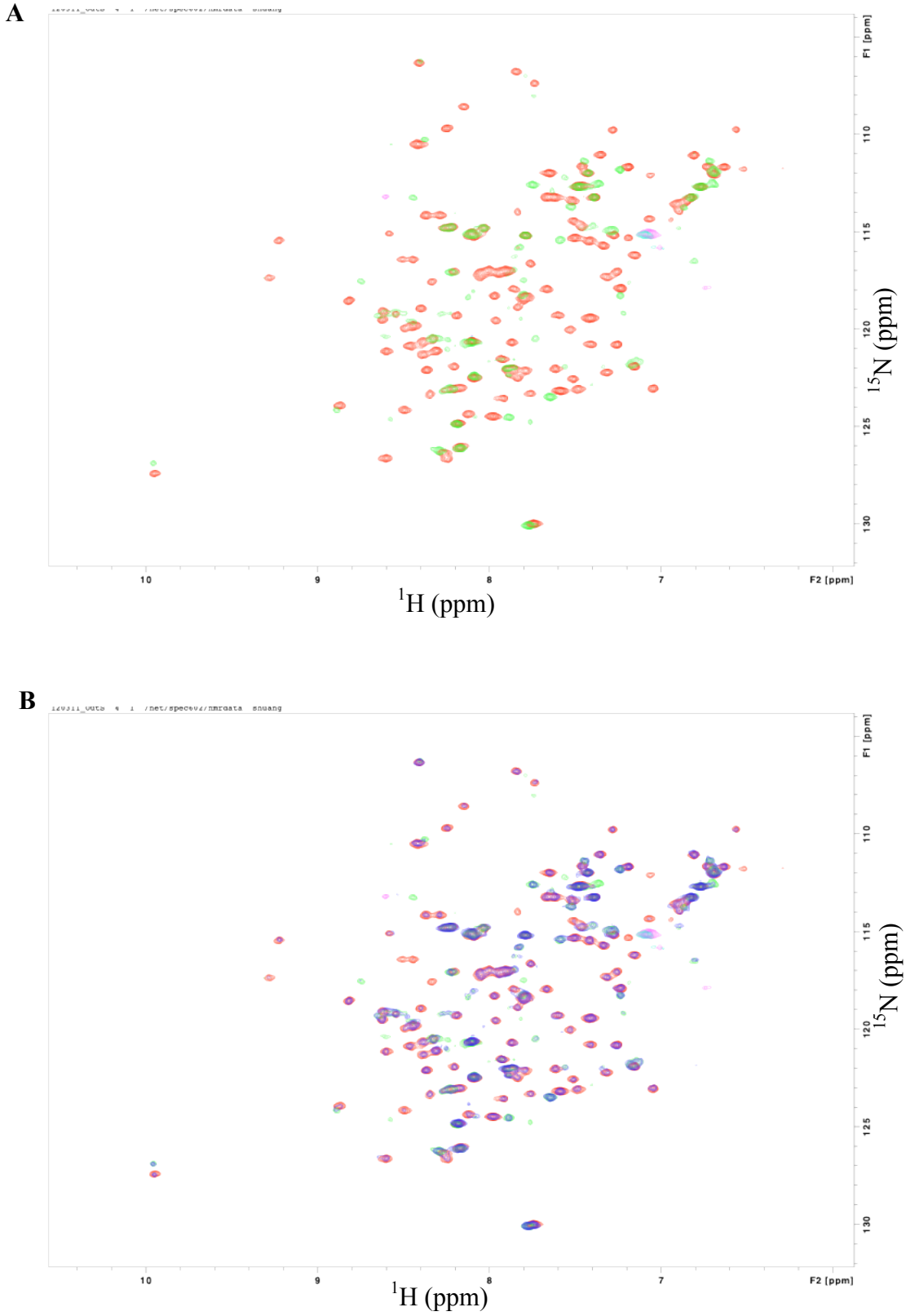


Figure 6.12 ^{1}H-^{15}N HSQC spectrum of OutS in the absence (red), with presence of 0.5 equivalents of OutD-Cter (blue) and 1.3 equivalents OutD-Cter (green). ^{15}N OutS was at 25 uM. Spectra were acquired on Bruker 600MHz at 15^{o}C in 20mM Tris 150mM NaCl (pH 7.0) and 10% $^{2}H_2O$.

not interactwith OutS, OutS was titrated in Looking closely at Figure 6.13A, the E708 is

the last observable shifted peak. To confirm that the OutD-Cter-cleaved, does not interactwith OutS, OutS was titrated in Looking closely at Figure 6.13A, G688, G699 and N690 do not behaves as in slow to OutD-Cter-cleaved (Figure 6.13B). In the presence of two equivalent of OutS, which is sufficient to saturate the OutD-Cter to OutS bound state, no change in the spectra was observed. Therefore, the OutD-Cter-cleaved peptide does not interact with OutS. This highlights the last 20 residues of OutD, $OutD^{691-710}$, as those that bind OutS.

The J-pred secondary structure prediction predicts helix from residues 694 to 704. Therefore OutD-Cter is probably interacting with OutS with a helix formed between these residues.

6.8 $OutD^{692-708}$ interacting with OutS

The studies described above strongly suggest the $OutD^{691-710}$ peptide interacts with OutS. To confirm this an 18 residue peptide was purchased from Generon and tested in NMR titration experiments.

The CD experiment showed the 18 residues $OutD^{692-708}$ peptide was unstructured (Figure 6.14). This result is expected and consistent with the CD result for the OutD-Cter. The $OutD^{692-708}$ was subsequently used for NMR titration studies. $OutD^{692-708}$ was titrated against ^{15}N OutS. With addition of $OutD^{692-708}$, many peaks of on OutS spectra shifted in a slow exchange manner as previously seen for the OutD-Cter titration

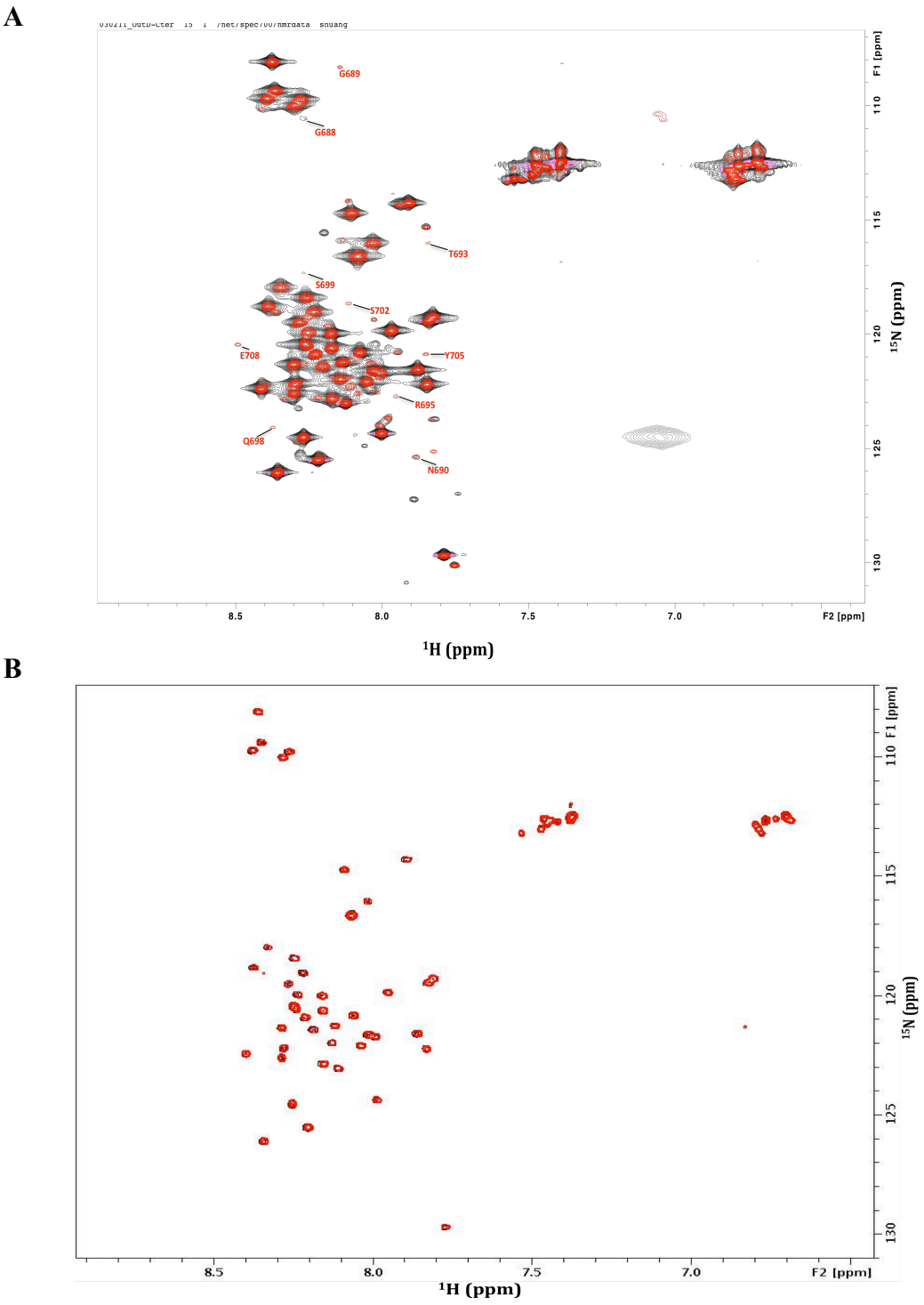


Figure 6.13 OutD-Cter* and OutD-Cter-cleaved titration with OutS. (A) 2D ^{1}H-^{15}N HSQC of ^{15}NOutD-Cter* (B) 2D ^{1}H-^{15}N HSQC of ^{15}NOutD-Cter-cleaved in the absence (black) and presence of two equivalent of OutS(red). ^{15}NOutD-Cter* was measured at 50mM and ^{15}NOutD-Cter-cleaved was measure at 25mM. Both spectrum acquired at15°C on Bruker 700MHz at 15^{0}C in 20mM Tris 150mM NaCl (pH 7.0) and 10% $^{2}H_2O$.

(Figure 6.15). Comparing the OutS HSQCs in the presence of OutD-Cter and OutD$^{692-708}$, the changes on the OutS spectra are very similar (Figure 6.15C) This confirms OutD residues 692-708 is the major region, and probably the only region, responsible for interacting with OutS.

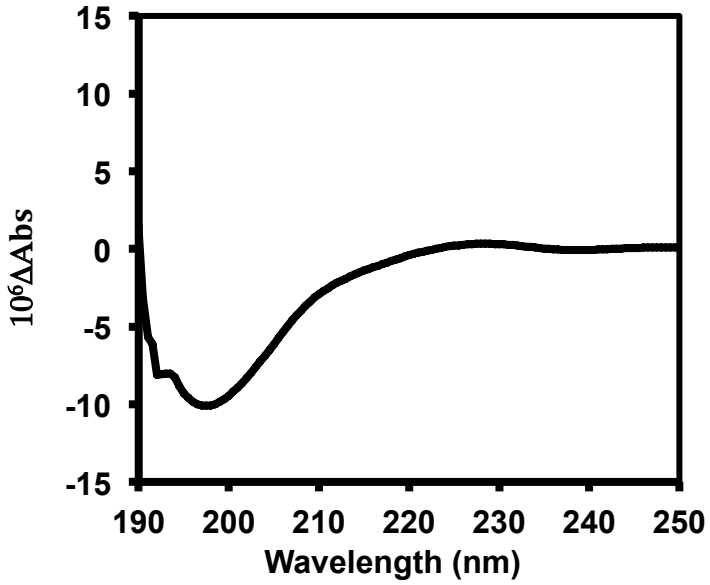
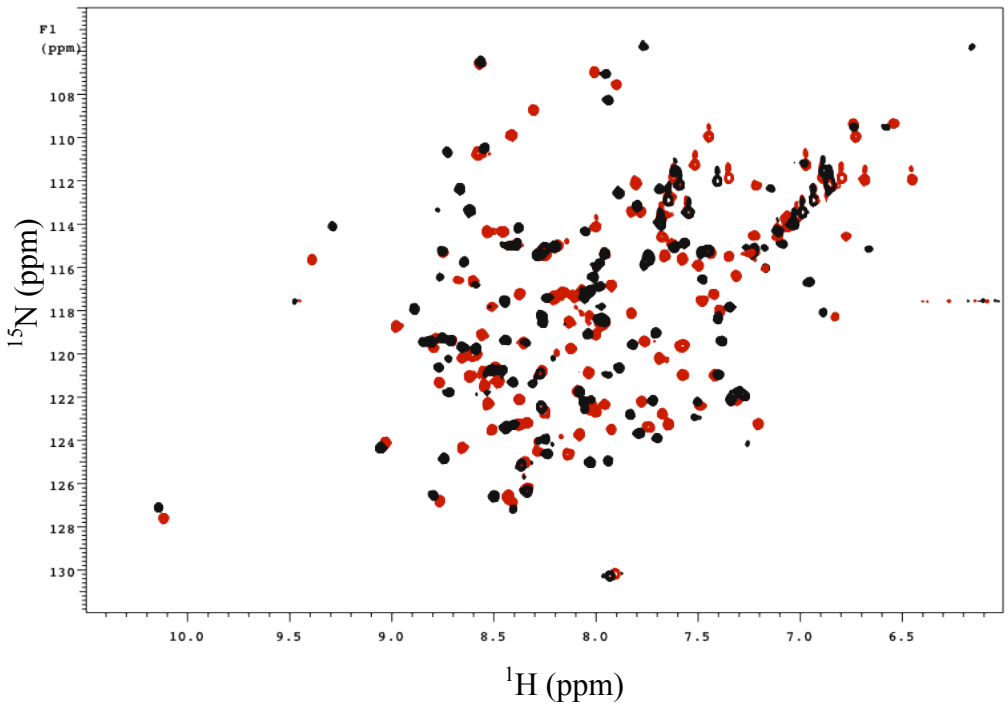
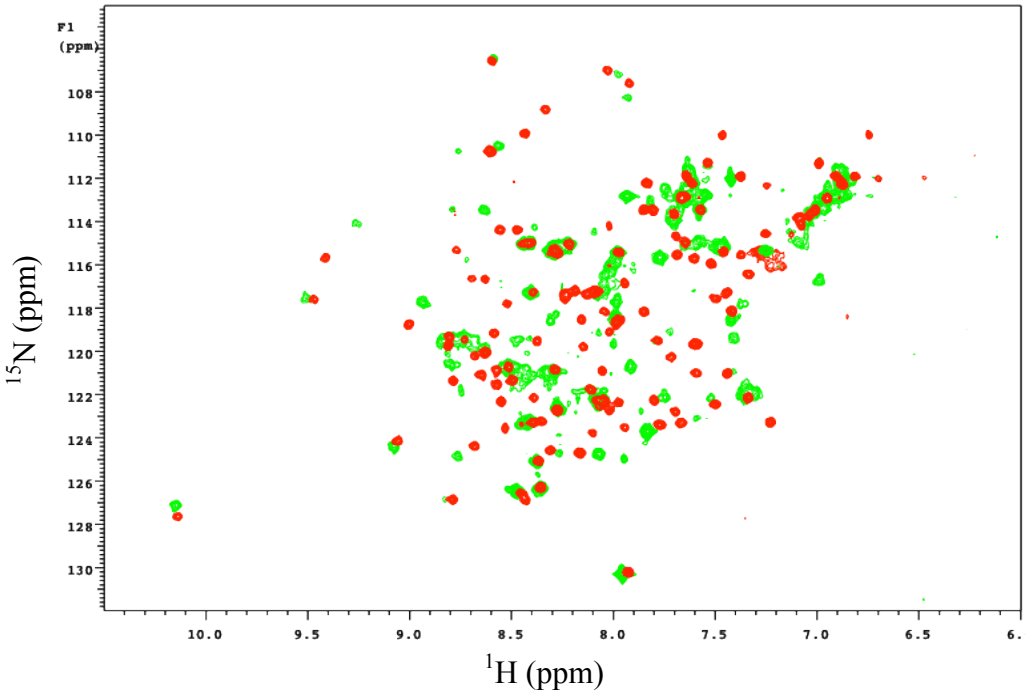


Figure 6.14 Far-UV CD spectra of OutD$^{692-708}$. Spectra was measured at 20^{O}C in 20mM Tris, 150mM NaCl pH7.0.

A



B



C

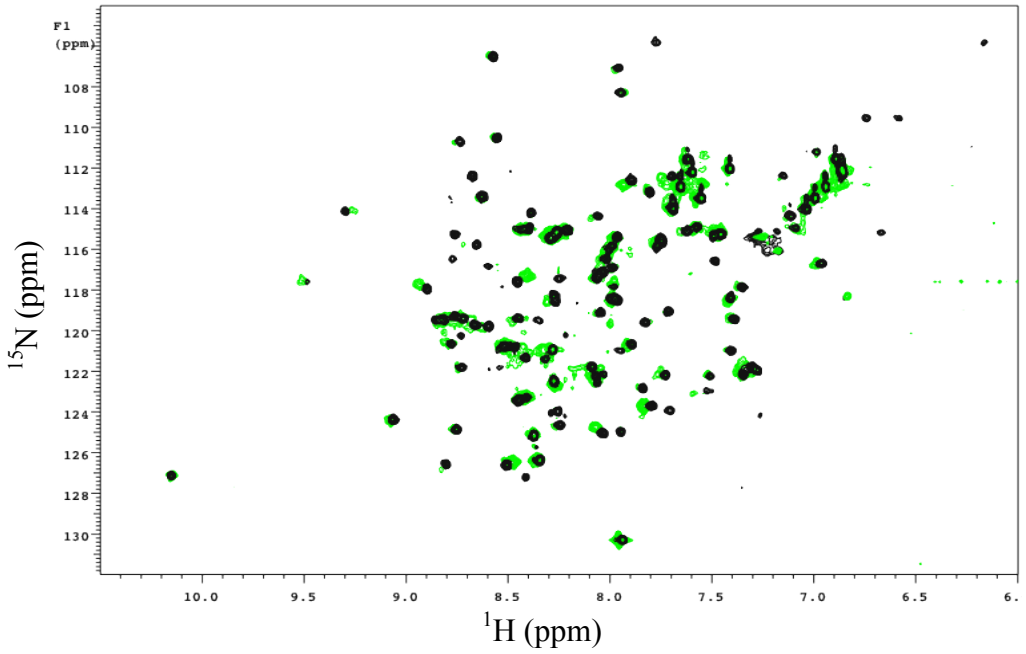


Figure 6.15 Comparing effect of OutD$^{692-708}$ and OutD-Cter on OutS. (A)^{1}H-^{15}N HSQC spectrum of OutS in the absence (red) and presence of OutD$^{692-708}$ (black). (B) ^{1}H-^{15}N HSQC spectrum of OutS in the absence (red) and presence of OutD-Cter (green). (C) Overlay of ^{1}H-^{15}N HSQC spectrum of OutS in the presence of OutD$^{692-708}$ (black) and ^{1}H-^{15}N HSQC spectrum of OutS in the presence of OutD-Cter (green). ^{15}N OutS was measured at 100mM, OutD$^{692-708}$ was measured at 100mM and OutD-Cter was measured at 130mM. The spectra were acquired at Varian 800MHz at 15^{o}C in 20mM Tris 150mM NaCl (pH 7.0) and 10% $^{2}H_2O$.

Since OutS has only one tryptophan and there was no tryptophan in $OutD^{691-708}$, fluorescence spectroscopy by measuring tryptophan emission is used to observing the interaction. Upon binding to the $OutD^{691-708}$ the tryptophan signal is quenched. The binding between OutS and $OutD^{691-708}$ appears to be very tight with dissociation constant (K_d) of 55.0±20nM, error was estimated by three independent measurments.

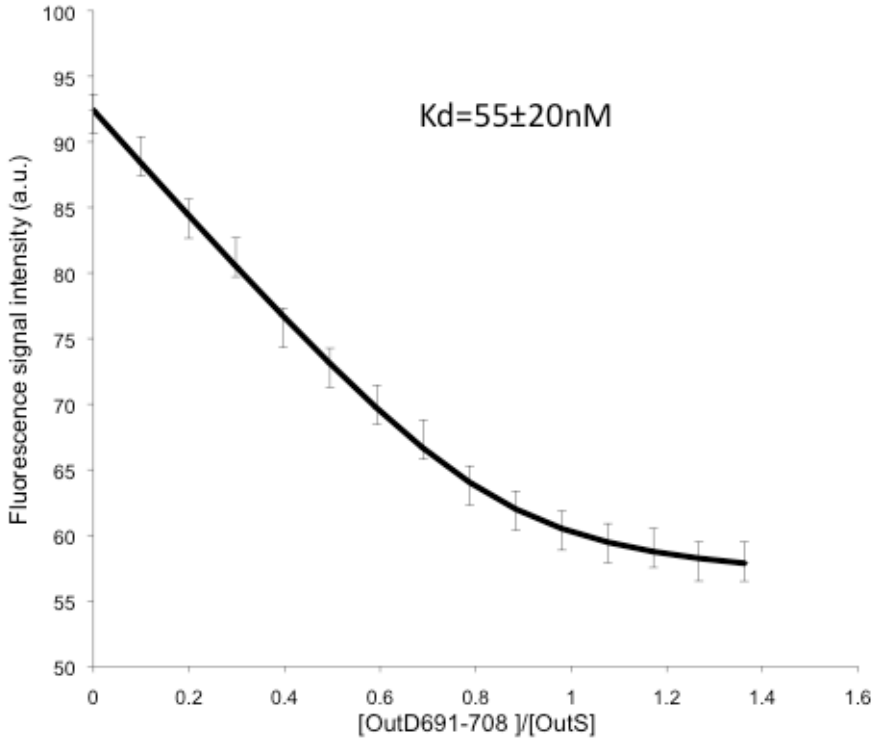


Figure 6.16 OutS and $OutD^{691-708}$ K_d determination by flouorescence spectroscopy. $OutD^{691-708}$ titration of OutS as monitored by tryptophan emission at 340nm. Spectra of 1μM OutS was measured with variable $OutD^{691-708}$ at 15°C in 20mM Tris, 150mM NaCl, pH 7.0. Standard deviation was calculated from three different measurments.

The OutS has a nested four α-helices structure (Figure 6.17A), revealed a concave surface formed by juxtaposition of helices 1, 3 and 4 and this structure is likely to be the binding site for the OutD-Cter. When OutS is compared with the P40 nucleoprotein (Rudolph, Kraus et al. 2003), the protein with most similar arrangement of alpha-helics, the corresponding site is occupied by helix 3 (Figure 6.17). This similarity suggests that

the concave surface formed by helices 1,3 and 4 of OutS bind a helical structure. Secondary structure prediction on OutD-Cter suggests that this region consists of three alpha-helices. Although the OutD-Cter HSQC spectrum indicates the OutD-Cter is lack of secondary structure, the peaks showed slow exchange during the titration experiment may form helices structure upon binding to the OutS. To localize the region of the interaction on OutD-Cter, the OutD-Cter backbone assignment was achieved. Residues corresponding to $OutD^{692-708}$ behave as if they are in a slow exchange regime during titration and were located to be the region of interaction. The CD results suggesting OutD-Cter forming helical structure when interacting with OutS. Although OutD-Cter was unstructured on its own, it is predicted to have helical structure and able to form helical structure. Therefore, OutD-Cter C-terminus predicted helical is predicted to interact with the concave surface of OutS in a manner similar to helix 3 of P40 nucleoprotein.

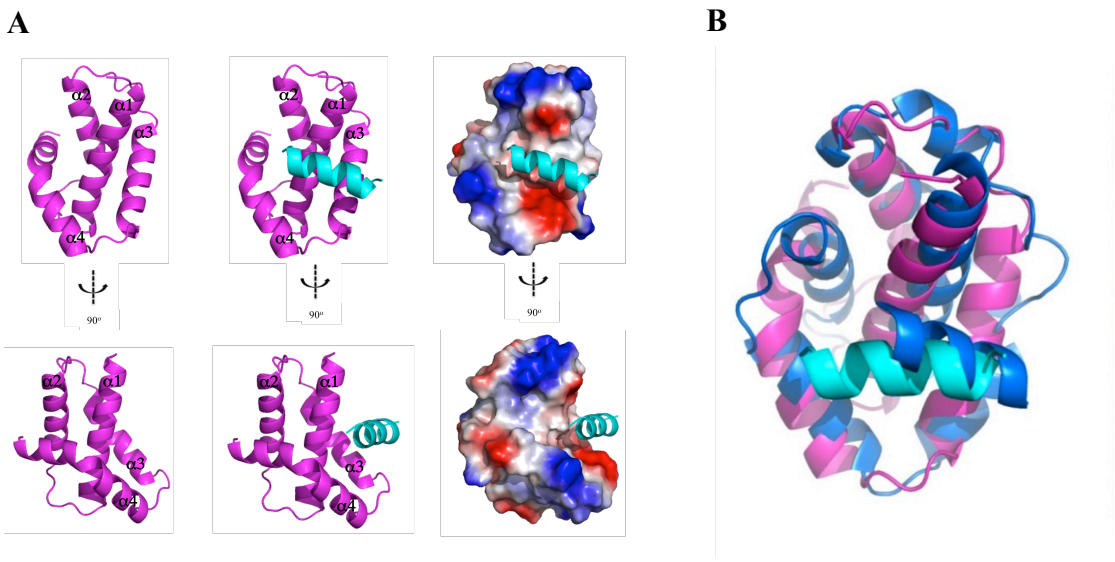


Figure 6.17 Proposed OutS OutD binding model (A) OutS structure (megenta) with proposed OutD peptid (cyan) binding model. (B) Superimposition of P40 nucleoprotein (blue, PDB:1N93), the OutS (magenta) and proposed OutD peptide (cyan). The P40 nucleoprotein has been simplified to those elements with equivalents in OutS, with the exception of P40 helix 3 closest to the viewer.

6.9 Summary and discussion

In this chapter, an 18 residue sequence on OutD, $OutD^{692-708}$, was found to be responsible for interacting with OutS. Upon binding to OutS, unstructured $OutD^{692-708}$ forms a more helical structure.

The pilotin lipidation is essential for its function. In type three secretion system, an unlipidated form of the YscW protein was not functional, although it still interacted with the secretin and caused mislocalization of YscC even in the presence of wild-type YscW (Burghout, Beckers et al. 2004). Also the type III pilotin, MxiM binding site is occupied by it's own lipid acyl chain in the absence of MxiD (Lario, Pfuetzner et al. 2005). The binding of lipid seems playing a regulatory role, that only the secretin binding MxiM has the acyl chain presented to the LolA for transportation. This ensures only the secreting associated pilotin is transported. The binding between OutS and $OutD^{692-708}$ is also dominated by hydrophobic interactions. Therefore, OutS binding surface is predicted to occupied by its own acyl lipid chain in the absence of OutD. Upon binding OutD C-terminus, OutS acyl lipid chain is expelled and presented to LolA for Lol translocation to the outer membrane (Figure 6.24).

GspS is only identified in a few species (Shevchik, Robert-Baudouy et al. 1997; Nouwen, Ranson et al. 1999). But it maybe more wide spread. When known GspS sequences are used in BLAST searches, many uncharacterized proteins with similar size, from different species, are found. These hits are good candidates for pilotins. As suggested from examination of the type III pilotin, MxiM, mechanism, it is likely that

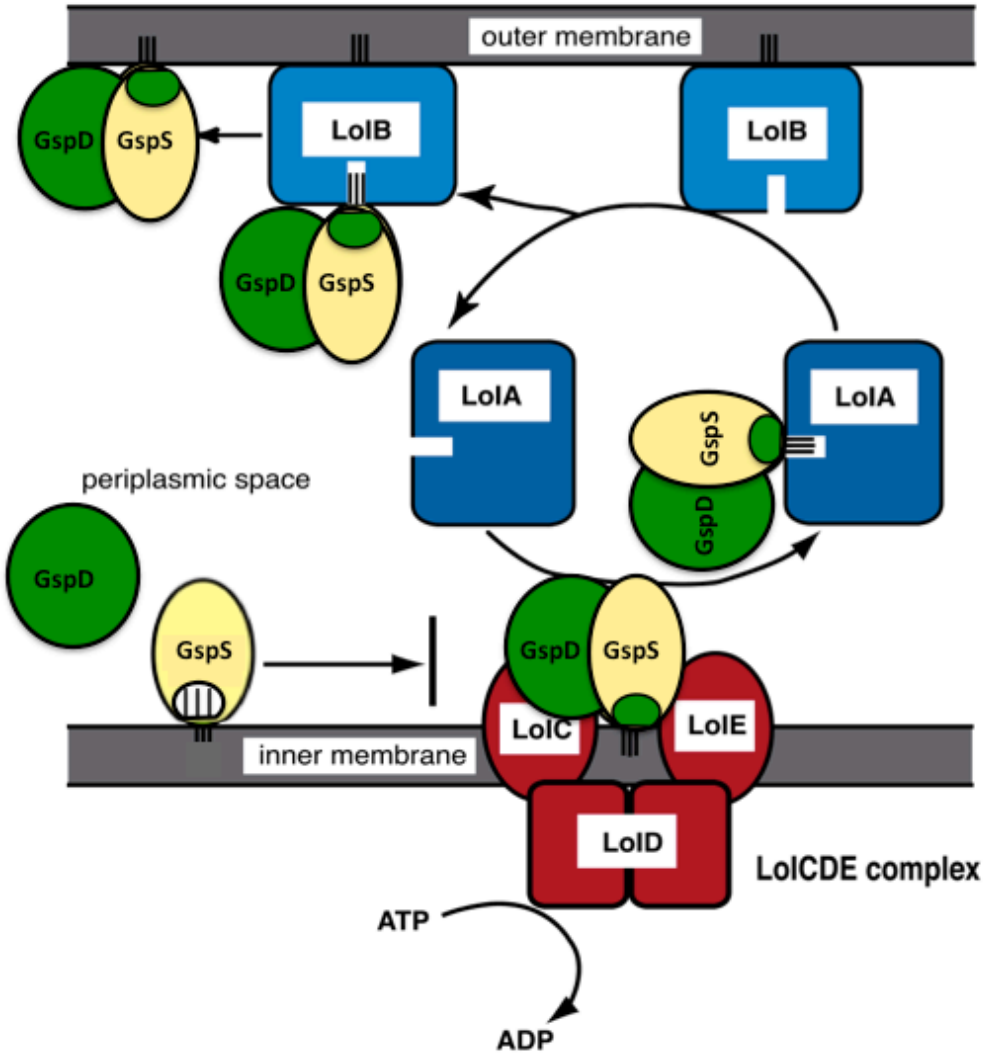


Figure 6.18 A model of tranlocation of GspS and GspD to the outer membrane via Lol pathway (Tokuda and Matsuyama 2004).

GspS binds an acyl lipid in the absence of GspD. There is currently no experimental evidence to support this. It would be interesting to test this. If GspS does bind lipid, stabilizing the lipid binding could be a possibility to inhibit T2SS related diseases. The alternative, of course, is to inhibit interaction with secretin.

In this work, an 18 residues sequence from GspD was identified that binds OutS. Although a predicted model is proposed, it would be useful to study the interaction in molecular detail, that is to solve the OutS/18 residues complex structure by NMR or

crystallography. GspS could be a very good drug target for several T2SS related diseases as blocking its binding to LolA or secretin could completely abolish T2SS assembly and stop bacteria secreting toxins.

Chapter 7 Overall Conclusions

In the past decade, there have been great advances in understanding the molecular structure of the T2SS. The structure of a great many single soluble components and domains have been solved. But there still limited knowledge of how the components assemble to form a functional system. In this Thesis I have focussed on the elucidating the molecular details of the interaction of the T2SS outer membrane secretin, GspD, with the inner-membrane protein GspC and the pilotin GspS.

In this work, I have solved the solution structure of GspC-HR by using NMR spectroscopy, which revealed it to have a β-sandwich fold. The structure of OutC-HRF3 shares a very similar fold as a T4P lipoprotein PilP. PilP shares many functional roles with GspC-HR, which include locating on the inner membrane, extending into periplasm and interacting with an outer membrane secretin (Golovanov, Balasingham et al. 2006). Both GspC and PilP form inner-membrane subcomplexes (GspC/M/L and PilP/N/M) with other components of their secretion systems(Tammam, Sampaleanu et al. 2011). The recently released PilM-PilN complex structure (PDB:2YCH) has a similar ferredoxin-like fold as seen in the GspL-GspM complex (Karuppiah and Derrick ; Abendroth, Rice et al. 2004; Abendroth, Kreger et al. 2009). Even though they have only 18% sequence similarity, GspC-HR and PilP should be considered as homologus between T2SS and T4P. I also show that, GspC-HRF3 interacts with the N0 domain of the secretin (GspD) subunit. Currently, there is no molecular detail on PilP and PilQ interaction, but it is very likely that PilP interacts with secretin PilQ in a very similar manner through interaction with PilQ N0 domain (Tammam, Sampaleanu et al. 2011). In the map from EM reconstruction of the PilP and PilQ complex (Balasingham, Collins et al. 2007), PilP interacts with PilQ in the periplasm and this is where the N0 domain

is anticipated to reside.

T2SS and T4P have long-recognized similarities. Many of T2SS components have homologoues counterparts in the T4P (Peabody, Chung et al. 2003). These include (1) the prepilin peptidase/N- methyltransferase (GspO/PilD), (2) the ATPase (GspE/PilB), (3) the secretin (GspD/PilQ) and (4) the multispanning transmembrane (TM) protein (GspF/PilC), (5) inner-membrane subcomplex proteins (GspL and M/ PilN and M). In this work another structure of GspC-HR was solved by using NMR spectrascopy, which revealed it it a homologue of PilP. Out of all the homologues, only two constituents, the ATPase and the TM protein were identified from sequence and phylogenetic analyses. Although many other protein constituents are probably homologous, extensive sequence divergence during evolution clouds this homology.

The structural homology between the T2SS and T4P doesnot extend to the secretin's pilotin of two secretion systems. The T2SS GspS is a much smaller protein in size than the T4P PilW. Recently solved GspS structure revealed a new fold different to PilW structure. In this work, the interaction between secretin and pilotin is studied by a series of biophysical techniques. The region of GspS binding site on secretin is shown to be limited to the C-terminal 18 residues. The unstructured GspD C-terminal 18 residues are also shown to form a helical structure upon binding to pilotin. It interesting to mention the T3SS pilotin MxiM, with a different structure also binds to the C-terminal 18 residues of the secretin in a similar manner where the unstructured peptide of secretin forming helical structure upon binging to MxiM (Okon, Moraes et al. 2008). Although the T2SS and T3SS pilotins do not show structural homology, they have clear functional and mechanistic similarities.

References

Abendroth, J., A. C. Kreger, et al. (2009). "The dimer formed by the periplasmic domain of EpsL from the Type 2 Secretion System of Vibrio parahaemolyticus." Journal of Structural Biology 168(2): 313-322.

Abendroth, J., D. D. Mitchell, et al. (2009). "The three-dimensional structure of the cytoplasmic domains of EpsF from the type 2 secretion system of Vibrio cholerae." Journal of Structural Biology 166(3): 303-315.

Abendroth, J., P. Murphy, et al. (2005). "The X-ray Structure of the Type II Secretion System Complex Formed by the N-terminal Domain of EpsE and the Cytoplasmic Domain of EpsL of Vibrio cholerae." Journal of Molecular Biology 348(4): 845-855.

Abendroth, J., A. E. Rice, et al. (2004). "The Crystal Structure of the Periplasmic Domain of the Type II Secretion System Protein EpsM From Vibrio cholerae: The Simplest Version of the Ferredoxin Fold." Journal of Molecular Biology 338(3): 585-596.

Balasingham, S. V., R. F. Collins, et al. (2007). "Interactions between the Lipoprotein PilP and the Secretin PilQ in Neisseria meningitidis." J Bacteriol. 189(15): 5716-5727.

Bax, A. and A. Grishaev (2005). "Weak alignment NMR: a hawk-eyed view of biomolecular structure." Curr Opin Struct Biol 15(5): 563-570.

Berjanskii, M. V. and D. S. Wishart (2005). "A Simple Method To Predict Protein Flexibility Using Secondary Chemical Shifts." Journal of the American Chemical Society 127(43): 14970-14971.

Bleves, S., M. Gerard-Vincent, et al. (1999). "Structure-function analysis of XcpP, a component involved in general secretory pathway-dependent protein secretion in Pseudomonas aeruginosa." J Bacteriol 181(13): 4012-4019.

Bouley, J., G. Condemine, et al. (2001). "The PDZ domain of OutC and the N-terminal region of OutD determine the secretion specificity of the type II out pathway of Erwinia chrysanthemi." J Mol Biol 308(2): 205-219.

Burghout, P., F. Beckers, et al. (2004). "Role of the Pilot Protein YscW in the Biogenesis of the YscC Secretin in Yersinia enterocolitica." J. Bacteriol. 186(16): 5366-5375.

Camberg, J. L. and M. Sandkvist (2005). "Molecular Analysis of the Vibrio cholerae Type II Secretion ATPase EpsE." J. Bacteriol. 187(1): 249-256.

Chami, M., I. Guilvout, et al. (2005). "Structural Insights into the Secretin PulD and Its Trypsin-resistant Core." Journal of Biological Chemistry 280(45): 37732-37741.

Cianciotto, N. (2005). "Type II secretion: a protein secretion system for all seasons." Trends in Microbiology 13(12): 581-588.

Collin, S., I. Guilvout, et al. (2011). "Sorting of an integral outer membrane protein via the lipoprotein-specific Lol pathway and a dedicated lipoprotein pilotin." Molecular Microbiology 80(3): 655-665.

Condemine, G. and V. E. Shevchik (2000). "Overproduction of the secretin OutD suppresses the secretion defect of an Erwinia chrysanthemi outB mutant." Microbiology 146(3): 639-647.

Corbett, M., S. Virtue, et al. (2005). "Identification of a New Quorum-Sensing-Controlled Virulence Factor in Erwinia carotovora subsp. atroseptica Secreted via the Type II Targeting Pathway." Molecular Plant-Microbe Interactions 18(4): 334-342.

Cordier, F. and S. Grzesiek (1999). "Direct Observation of Hydrogen Bonds in Proteins by Interresidue 3hJNC,Äò Scalar Couplings." Journal of the American Chemical Society 121(7): 1601-1602.

Crago, A. M. and V. Koronakis (1998). "Salmonella InvG forms a ring-like multimer that requires the InvH lipoprotein for outer membrane localization." Molecular Microbiology 30(1): 47-56.

d'Enfert C, Ryter A, et al. (1987). "Cloning and expression in Escherichia coli of the Klebsiella pneumoniae genes for production, surface localization and secretion of the lipoprotein pullulanase." EMBO J. 6(11): 3531-3538.

Daefler, S., I. Guilvout, et al. (1997). "The C-terminal domain of the secretin PulD contains the binding site for its cognate chaperone, PulS, and confers PulS dependence on pIVf1 function." Molecular Microbiology 24(3): 465-475.

Davis, B. M., E. H. Lawson, et al. (2000). "Convergence of the Secretory Pathways for Cholera Toxin and the Filamentous Phage, CTXœï." Science 288(5464): 333-335.

Delaglio, F., S. Grzesiek, et al. (1995). "NMRPipe: a multidimensional spectral processing system based on UNIX pipes." J Biomol NMR 6(3): 277-293.

Dominguez, C., R. Boelens, et al. (2003). "HADDOCK: a protein-protein docking approach based on biochemical or biophysical information." J Am Chem Soc. 125(7): 1731-1737.

Douet, V., L. Loiseau, et al. (2004). "Systematic analysis, by the yeast two-hybrid, of protein interaction between components of the type II secretory machinery of Erwinia chrysanthemi." Research in Microbiology 155(2): 71-75.

Drake, S. L., S. A. Sandstedt, et al. (1997). "PilP, a pilus biogenesis lipoprotein in Neisseria gonorrhoeae, affects expression of PilQ as a high-molecular-mass multimer." Mol Microbiol 23(4): 657-668.

Driessen, A. (1992). "Precursor protein translocation by the Escherichia coli translocase is directed by the protonmotive force." EMBO J 11(3): 847-853.

Durand, E., D. Verger, et al. (2009). "Structural Biology of Bacterial Secretion Systems in Gram-Negative Pathogens- Potential for New Drug Targets " Infectious Disorders - Drug Targets(9): 518-547.

Fries, M., J. Ihrig, et al. (2007). "Molecular basis of the activity of the phytopathogen pectin methylesterase." EMBO J 26(17): 3879-3887.

Genin, S. and C. A. Boucher (1994). "A superfamily of proteins involved in different secretion pathways in gram-negative bacteria: modular structure and specificity of the N-terminal domain." Molecular and General Genetics MGG 243(1): 112-118.

Golovanov, A. P., S. Balasingham, et al. (2006). "The Solution Structure of a Domain from the Neisseria meningitidis Lipoprotein PilP Reveals a New [beta]-Sandwich Fold." Journal of Molecular Biology 364(2): 186-195.

Goo, E., Y. Kang, et al. (2010). "Proteomic Analysis of Quorum Sensing-Dependent Proteins in Burkholderia glumae." Journal of Proteome Research 9(6): 3184-3199.

Guilvout, I., M. Chami, et al. (2006). "Bacterial outer membrane secretin PulD assembles and inserts into the inner membrane in the absence of its pilotin." EMBO J 25(22): 5241-5249.

Guilvout, I., N. N. Nickerson, et al. "Multimerization-defective variants of dodecameric secretin PulD." Research in Microbiology 162(2): 180-190.

Hardie, K. R., S. Lory, et al. (1996). "Insertion of an outer membrane protein in Escherichia coli requires a chaperone-like protein." EMBO J 15(5): 978-988.

Hardie, K. R., A. Seydel, et al. (1996). "The secretin-specific, chaperone-like protein of the general secretory pathway: separation of proteolytic protection and piloting functions." Mol Microbiol 22(5): 967-976.

Hayashi, S. and H. C. Wu (1990). "Lipoproteins in bacteria." Journal of Bioenergetics and Biomembranes 22(3): 451-471.

He, S. Y., M. Lindeberg, et al. (1991). "Cloned Erwinia chrysanthemi out genes enable Escherichia coli to selectively secrete a diverse family of heterologous proteins to its milieu." Proc Natl Acad Sci U S A 88(3): 1079-1083.

Hobot, J. A., E. Carlemalm, et al. (1984). "Periplasmic gel: new concept resulting from the reinvestigation of bacterial cell envelope ultrastructure by new methods." J Bacteriol. 160(1): 143-152.

Holm, L. and P. Rosenstrom (2010). "Dali server: conservation mapping in 3D." Nucleic Acids Res 38(Web Server issue): W545-549.

Jaravine, V. A., A. V. Zhuravleva, et al. (2008). "Hyperdimensional NMR Spectroscopy with Nonlinear Sampling." Journal of the American Chemical Society 130(12): 3927-3936.

Karuppiah, V. and J. P. Derrick "Structure of the PilM-PilN Inner Membrane Type IV Pilus Biogenesis Complex from Thermus thermophilus." Journal of Biological Chemistry 286(27): 24434-24442.

Kay LE, T. D., Bax A. (1989). "Backbone dynamics of proteins as studied by 15N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease." Biochemistry. 14(28): 8972-8979.

Kim, K., J. Oh, et al. (2006). "Crystal structure of PilF: Functional implication in the type 4 pilus biogenesis in Pseudomonas aeruginosa." Biochemical and Biophysical Research Communications 340(4): 1028-1038.

Koo J, T. S., Ku SY, Sampaleanu LM, Burrows LL, Howell PL. (2008). "PilF Is an Outer Membrane Lipoprotein Required for Multimerization and Localization of the Pseudomonas aeruginosa Type IV Pilus Secretin." J Bacteriol. 190(21): 6961-6969.

Korotkov, K. V., T. Gonen, et al. (2011). "Secretins: dynamic channels for protein transport across membranes." Trends in Biochemical Sciences.

Korotkov, K. V. and W. G. J. Hol (2008). "Structure of the GspK-GspI-GspJ complex from the enterotoxigenic Escherichia coli type 2 secretion system." Nat Struct Mol Biol 15(5): 462-468.

Korotkov, K. V., T. L. Johnson, et al. (2011). "Structural and Functional Studies on the Interaction of GspC and GspD in the Type II Secretion System." PLoS Pathog 7(9): e1002228.

Korotkov, K. V., B. Krumm, et al. (2006). "Structural and functional studies of EpsC, a crucial component of the type 2 secretion system from Vibrio cholerae." J Mol Biol 363(2): 311-321.

Korotkov, K. V., E. Pardon, et al. (2009). "Crystal structure of the N-terminal domain of the secretin GspD from ETEC determined with the assistance of a nanobody." Structure 17(2): 255-265.

Koster, M., W. Bitter, et al. (1997). "The outer membrane component, YscC, of the Yop secretion machinery of Yersinia enterocolitica forms a ring-shaped multimeric complex." Molecular Microbiology 26(4): 789-797.

Lario, P. I., R. A. Pfuetzner, et al. (2005). "Structure and biochemical analysis of a secretin pilot protein." EMBO J 24(6): 1111-1121.

Lee, H.-M., J.-R. Chen, et al. (2004). "Functional Dissection of the XpsN (GspC) Protein of the Xanthomonas campestris pv. campestris Type II Secretion Machinery." J. Bacteriol. 186(10): 2946-2955.

Lee, H.-M., K.-C. Wang, et al. (2000). "Association of the Cytoplasmic Membrane Protein XpsN with the Outer Membrane Protein XpsD in the Type II Protein Secretion Apparatus of Xanthomonas campestris pv. Campestris." J. Bacteriol. 182(6): 1549-1557.

Lindeberg, M., G. P. C. Salmond, et al. (1996). "Complementation of deletion mutations in a cloned functional cluster of Erwinia chrysanthemi out genes with Erwinia carotovora out homologues reveals OutC and OutD as candidate gatekeepers of species-specific secretion of proteins via the type II pathway." Molecular Microbiology 20(1): 175-190.

Linderoth, N. A., M. N. Simon, et al. (1997). "The Filamentous Phage pIV Multimer Visualized by Scanning Transmission Electron Microscopy." Science 278(5343): 1635-1638.

Linge, J. P., M. Habeck, et al. (2003). "ARIA: automated NOE assignment and NMR structure calculation." Bioinformatics 19(2): 315-316.

Liwang, A. C. and A. Bax (1997). "Solution NMR characterization of hydrogen bonds in a protein by indirect measurement of deuterium quadrupole couplings." J Magn Reson 127(1): 54-64.

Login, F. H., M. Fries, et al. (2010). "A 20-residue peptide of the inner membrane protein OutC mediates interaction with two distinct sites of the outer membrane secretin OutD and is essential for the functional type II secretion system in Erwinia chrysanthemi." Mol Microbiol 76(4): 944-955.

Login, F. H. and V. E. Shevchik (2006). "The single transmembrane segment drives self-assembly of OutC and the formation of a functional type II secretion system in Erwinia chrysanthemi." J Biol Chem 281(44): 33152-33162.

Martin, S. R. and M. J. Schilstra (2008). "Circular Dichroism and Its Application to the Study of Biomolecules." METHODS IN CELL BIOLOGY 84: 263-293.

Michel, G. r., S. Bleves, et al. (1998). "Mutual stabilization of the XcpZ and XcpY components of the secretory apparatus in Pseudomonas aeruginosa." Microbiology 144(12): 3379-3386.

Miller, M. B. and B. L. Bassler (2001). "Quorum sensing in bacteria " Annual Review of Microbiology 55: 165-199.

Mühlradt, P. and J. Golecki (1975). "Asymmetrical distribution and artifactual reorientation of lipopolysaccharide in the outer membrane bilayer of Salmonella typhimurium." Eur J Biochem. 51(2): 343-352.

Nikaido, H. (2003). "Molecular Basis of Bacterial Outer Membrane Permeability Revisited." Microbiol. Mol. Biol. Rev. 67(4): 593-656.

Nouwen, N., N. Ranson, et al. (1999). "Secretin PulD: Association with pilot PulS, structure, and ion-conducting channel formation." Proceedings of the National Academy of Sciences of the United States of America 96(14): 8173-8177.

Nouwen, N., H. Stahlberg, et al. (2000). "Domain structure of secretin PulD revealed by limited proteolysis and electron microscopy." EMBO J 19(10): 2229-2236.

Okon, M., T. F. Moraes, et al. (2008). "Structural Characterization of the Type-III Pilot-Secretin Complex from Shigella flexneri." Structure (London, England : 1993) 16(10): 1544-1554.

Peabody, C. R., Y. J. Chung, et al. (2003). "Type II protein secretion and its relationship to bacterial type IV pili and archaeal flagella." Microbiology 149(11): 3051-3072.

Pellecchia, M., P. Sebbel, et al. (1999). "Pilus chaperone FimC-adhesin FimH interactions mapped by TROSY-NMR." Nat Struct Mol Biol 6(4): 336-339.

Perombelon, M. C. (1991). "The Prokaryotes." 2899-2921.

Piddock, L. J. V. (2006). "Multidrug-resistance efflux pumps ? not just for resistance." Nat Rev Micro 4(8): 629-636.

Possot, O. M., G. Vignon, et al. (2000). "Multiple Interactions between Pullulanase Secreton Components Involved in Stabilization and Cytoplasmic Membrane Association of PulE." J. Bacteriol. 182(8): 2142-2152.

Pugsley, A. P. (1993). "The complete general secretory pathway in gram-negative bacteria." Microbiology and Molecular Biology Reviews 57(1): 50-108.

Ramachandran, G. N., C. Ramakrishnan, et al. (1963). "Stereochemistry of polypeptide chain configurations." Journal of Molecular Biology 7(1): 95-99.

Reichow, S. L., K. V. Korotkov, et al. (2010). "Structure of the cholera toxin secretion channel in its closed state." Nat Struct Mol Biol 17(10): 1226-1232.

Rückert, M. and G. Otting (2000). "Alignment of Biological Macromolecules in Novel Nonionic Liquid Crystalline Media for NMR Experiments." J. Am. Chem. Soc. 122(32): 7793-7797.

Rudolph, M. G., I. Kraus, et al. (2003). "Crystal Structure of the Borna Disease Virus Nucleoprotein." Structure 11(10): 1219-1226.

Sambrook, J. (2001). "Molecular Cloning: A Laboratory Manual 3rd Ed." Cold Spring Harbor Laboratory Press. Cold Spring Harbor, NY.

Sandkvist, M. (2001). "Biology of type II secretion." Molecular Microbiology 40(2): 271-283.

Sandkvist, M. (2001). "Type II Secretion and Pathogenesis." Infection and Immunity 69(6): 3523-3535.

Sandkvist, M., M. Bagdasarian, et al. (1995). "Interaction between the autokinase EpsE and EpsL in the cytoplasmic membrane is required for extracellular secretion in Vibrio cholerae." EMBO J. 14(18): 1664-1673.

Sandkvist, M., J. M. Keith, et al. (2000). "Two Regions of EpsL Involved in Species-Specific Protein-Protein Interactions with EpsE and EpsM of the General Secretion Pathway in Vibrio cholerae." J. Bacteriol. 182(3): 742-748.

Schuch, R. and A. T. Maurelli (2001). "MxiM and MxiJ, Base Elements of the Mxi-Spa Type III Secretion System of Shigella, Interact with and Stabilize the MxiD Secretin in the Cell Envelope." J. Bacteriol. 183(24): 6991-6998.

Shen, Y. and A. Bax (2007). "Protein backbone chemical shifts predicted from searching a database for torsion angle and sequence homology." Journal of Biomolecular NMR 38(4): 289-302.

Shen Y, D. F., Cornilescu G, Bax A. (2009). "TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts." J Biomol NMR. 44(4): 213-223.

Shevchik, V. E. and G. Condemine (1998). "Functional characterization of the Erwinia chrysanthemi OutS protein, an element of a type II secretion system." Microbiology 144 (Pt 11): 3219-3228.

Shevchik, V. E., J. Robert-Baudouy, et al. (1997). "Specific interaction between OutD, an Erwinia chrysanthemi outer membrane protein of the general secretory pathway, and secreted proteins." EMBO J 16(11): 3007-3016.

Spreter, T., C. K. Yip, et al. (2009). "A conserved structural motif mediates formation of the periplasmic rings in the type III secretion system." Nat Struct Mol Biol 16(5): 468-476.

Takeda, K., H. Miyatake, et al. (2003). "Crystal structures of bacterial lipoprotein localization factors, LolA and LolB." EMBO J 22(13): 3199-3209.

Tammam, S., L. M. Sampaleanu, et al. (2011). "Characterization of the PilN, PilO and PilP type IVa pilus subcomplex." Molecular Microbiology 82(6): 1496-1514.

Tauschek, M., R. J. Gorrell, et al. (2002). "Identification of a protein secretory pathway for the secretion of heat-labile enterotoxin by an enterotoxigenic strain of Escherichia coli." Proceedings of the National Academy of Sciences 99(10): 7066-7071.

Thanassi, D. G. and S. J. Hultgren (2000). "Multiple pathways allow protein secretion across the bacterial outer membrane." Current Opinion in Cell Biology 12(4): 420-430.

Tokuda, H. (2009). "Biogenesis of Outer Membranes in Gram-Negative Bacteria." Biosci. Biotechnol. Biochem. 73: 465-473.

Tokuda, H. and S.-i. Matsuyama (2004). "Sorting of lipoproteins to the outer membrane in E. coli." Biochimica et Biophysica Acta (BBA) - Molecular Cell Research 1693(1): 5-13.

Tseng, T., B. Tyler, et al. (2009). "Protein secretion systems in bacterial-host associations, and their description in the Gene Ontology." BMC Microbiol. 19(9).

Warrena, J. J. and P. B. Moore (2002). "A Maximum Likelihood Method for Determining DaPQ and R for Sets of Dipolar Coupling Data." Journal of Magnetic Resonance 149(2): 271-275.

Wishart, D. S. and B. D. Sykes (1994). "Chemical shifts as a tool for structure determination." Methods in enzymology 239: 363-392.

Wooldridge, K. (2009). "Bacterial secreted proteins: secretory mechanisms and role in pathogenesis." Caister Academic Press.

Yakushi, T., N. Yokota, et al. (1998). "LolA-dependent Release of a Lipid-modified Protein from the Inner Membrane of Escherichia coli Requires Nucleoside Triphosphate." Journal of Biological Chemistry 273(49): 32576-32581.

Yamagata, A. and J. A. Tainer (2007). "Hexameric structures of the archaeal secretion ATPase GspE and implications for a universal secretion mechanism." EMBO J 26(3): 878-890.

Yamaguchi, K., F. Yu, et al. (1988). "A single amino acid determinant of the membrane localization of lipoproteins in E. coli." Cell 53(3): 423-432.

Appendices

Appendix 1 Protein sequences used for assignment.

OutC-HRF3 (pET14b after thrombin cleavage) sequence:

```
         10          20          30          40          50          60
GSHMLEMAGA LDASQMSNLP PSTLNLSLTG VMAGDDDSRS IAIISKDNEQ FSRGVNEEVP

         70          80          90         100
GYNAKIVSIR PDRVVLQYQG RYEVLGLYSQ EDSGSDGVPG AQVR
```

OutD-N0 (pET20b) sequence:

```
         10          20          30          40          50          60
MAEFSASFKG TDIQEFINTV SKNLNKTVII DPTVRGTISV RSYDMMNEGQ YYQFFLSVLD

         70          80          90         100
VYGFSVVPMD NGVLKVIRSK DAKSSSVDKL AAALEHHHHH H
```

OutD-Cter (pGEX-6p-2 after PreScission protease cleavage) sequence

```
         10          20          30          40          50          60
GPLGSPNSRV DPGQFQEASI NKYRSFNNEQ QQQRGEGNGV LDNNTLRLSG GNTYTFRQVQ

         70
SSISDFYKPE GRHHHHHH
```

OutD$^{691-708}$ (synthesized peptide by Generon) sequence

```
         10
TYTFRQVQSS ISDFYKPE
```

OutS (pET20b) sequence

```
         10          20          30          40          50          60
VKNTASRSAA SVPANEQISQ LASLVAASKY LRVQCERSDL PDDGTILKTA VNVAVQKGWD

         70          80          90         100
TGRYQSLPQL SENLYQGLLK DGTPKATQCS SFNRTMTPFL DAMRTVR
```

Construct name	Primer sequence	Primer direction
OutC-HRF3 (pET 14b)	5’ CATATG CTCGAG ATGGC GGGCG CGCTG GATGC G 3’	Forward
	5’ GTTATT GCTCAGC TTACC GGACC TGCGC GCCAG G 3’	Reverse
OutC-HRF2 (pET 14b)	5’ CATATG CTCGAG ATGCA GCAGC CGGTC ACCCT 3’	Forward
	5’ GTTATT GCTCAGC TTACC GGACC TGCGC GCCAG G 3’	Reverse
OutC-HRF4 (pET 14b)	5’ CATATG CTCGAG ATGCA GCAGC CGGTC ACCCT 3’	Forward
	5’ GTTATT GCTCAGC TTAAC GCTGA CCATC CGGA 3’	Reverse
OutD-N0 (pET 20b)	5’ CGCCATATGATG TGGGCTGCCG AATTTTC 3’	Forward
	5’ GGCCAATAACGAATGAGGATCCTGACCG 3’	Reverse

Appendix 2 ARIA 1.2 run setup script for structure calculation.

<html>
<head>
<title>ARIA - start</title>
</head>
<body bgcolor=#ffffff>
<h2>Parameters for the start:</h2>

<h4><!-- ARIA -->
ARIA_DIR=/dms/prog/bin/aria1.2
ASPECTRUM_1=N15
ASPECTRUM_2=C13
HBOND_FILE=/dms/home/shuang/ARIA_TEST/Shuang_H_bond_correct5.tbl
HET1_1=1
HET1_2=1
HET1_3=N
HET1_4=N
HET1_5=N
HET2_1=N
HET2_2=N
HET2_3=N
HET2_4=N
HET2_5=N
PPMDHET1_1=0.75
PPMDHET1_2=0.75
PPMDPRO1_1=0.035
PPMDPRO1_2=0.035
PPMDPRO2_1=0.04
PPMDPRO2_2=0.04
PRO1_1=2
PRO1_2=3
PRO1_3=N
PRO1_4=N
PRO1_5=N
PRO2_1=3
PRO2_2=2
PRO2_3=N
PRO2_4=N
PRO2_5=N
PROJECT_DIR=/dms/home/shuang/ARIA_TEST
RDC1_FILE=/dms/home/shuang/ARIA_TEST/HRF3_rdc.tbl
RUN_NUMBER=43_rdc19_test32_5_22_88_18_100
SEQ_PDB_FILE=/dms/home/shuang/ARIA_TEST/HRF3.seq
TALOS_ERROR=20.0
TALOS_FACTOR=2.0
TALOS_FILE=/dms/home/shuang/HRF3_talos_plus/talos+_C3.tab
UNAMBIG_TBL=/dms/home/shuang/ARIA_TEST/HRF3_UNAMBIG_13_4.tbl
WHICH_PEAKS_1=all
WHICH_PEAKS_2=all
WHICH_PEAKS_3=all
WHICH_PEAKS_4=all
WHICH_PEAKS_5=all
WHICH_SETUP=XEASY
XPEAKS_1=/dms/home/shuang/SGU_ARIA/N15_no_diag_ReNo_Relax_D_C.peaks
XPEAKS_2=/dms/home/shuang/SGU_ARIA/C13_no_diag_ReNo_2_Relax22_88_C.peaks
XPROT_1=/dms/home/shuang/ARIA_TEST/HRF3_correct3_22-88.prot
XPROT_2=/dms/home/shuang/ARIA_TEST/HRF3_correct3_22-88.prot
submit_save=Save updated parameters
</h4><!-- ARIA -->
</body>
</html>.

Appendix 3 Hydrogen bond restrains for OutC-HRF3

```
assign (resid 45 and name HN )(resid 27 and name O ) 2.1 0.5 0.5
assign (resid 45 and name N  )(resid 27 and name O ) 3.0 0.5 0.5
assign (resid 76 and name HN )(resid 83 and name O ) 2.1 0.5 0.5
assign (resid 76 and name N  )(resid 83 and name O ) 3.0 0.5 0.5
assign (resid 32 and name HN )(resid 41 and name O ) 2.1 0.5 0.5
assign (resid 32 and name N  )(resid 41 and name O ) 3.0 0.5 0.5
assign (resid 78 and name HN )(resid 81 and name O ) 2.1 0.5 0.5
assign (resid 78 and name N  )(resid 81 and name O ) 3.0 0.5 0.5
assign (resid 85 and name HN )(resid 74 and name O ) 2.1 0.5 0.5
assign (resid 85 and name N  )(resid 74 and name O ) 3.0 0.5 0.5
assign (resid 74 and name HN )(resid 85 and name O ) 2.1 0.5 0.5
assign (resid 74 and name N  )(resid 85 and name O ) 3.0 0.5 0.5
assign (resid 70 and name HN )(resid 73 and name O ) 2.1 0.5 0.5
assign (resid 70 and name N  )(resid 73 and name O ) 3.0 0.5 0.5
assign (resid 75 and name HN )(resid 68 and name O ) 2.1 0.5 0.5
assign (resid 75 and name N  )(resid 68 and name O ) 3.0 0.5 0.5
assign (resid 41 and name HN )(resid 32 and name O ) 2.1 0.5 0.5
assign (resid 41 and name N  )(resid 32 and name O ) 3.0 0.5 0.5
assign (resid 34 and name HN )(resid 36 and name O ) 2.1 0.5 0.5
assign (resid 34 and name N  )(resid 36 and name O ) 3.0 0.5 0.5
assign (resid 44 and name HN )(resid 51 and name O ) 2.1 0.5 0.5
assign (resid 44 and name N  )(resid 51 and name O ) 3.0 0.5 0.5
assign (resid 66 and name HN )(resid 57 and name O ) 2.1 0.5 0.5
assign (resid 66 and name N  )(resid 57 and name O ) 3.0 0.5 0.5
assign (resid 43 and name HN )(resid 30 and name O ) 2.1 0.5 0.5
assign (resid 43 and name N  )(resid 30 and name O ) 3.0 0.5 0.5
assign (resid 53 and name HN )(resid 42 and name O ) 2.1 0.5 0.5
assign (resid 53 and name N  )(resid 42 and name O ) 3.0 0.5 0.5
```

Appendix 4 Residual dipolar coupling restrains for OutC-HRF3

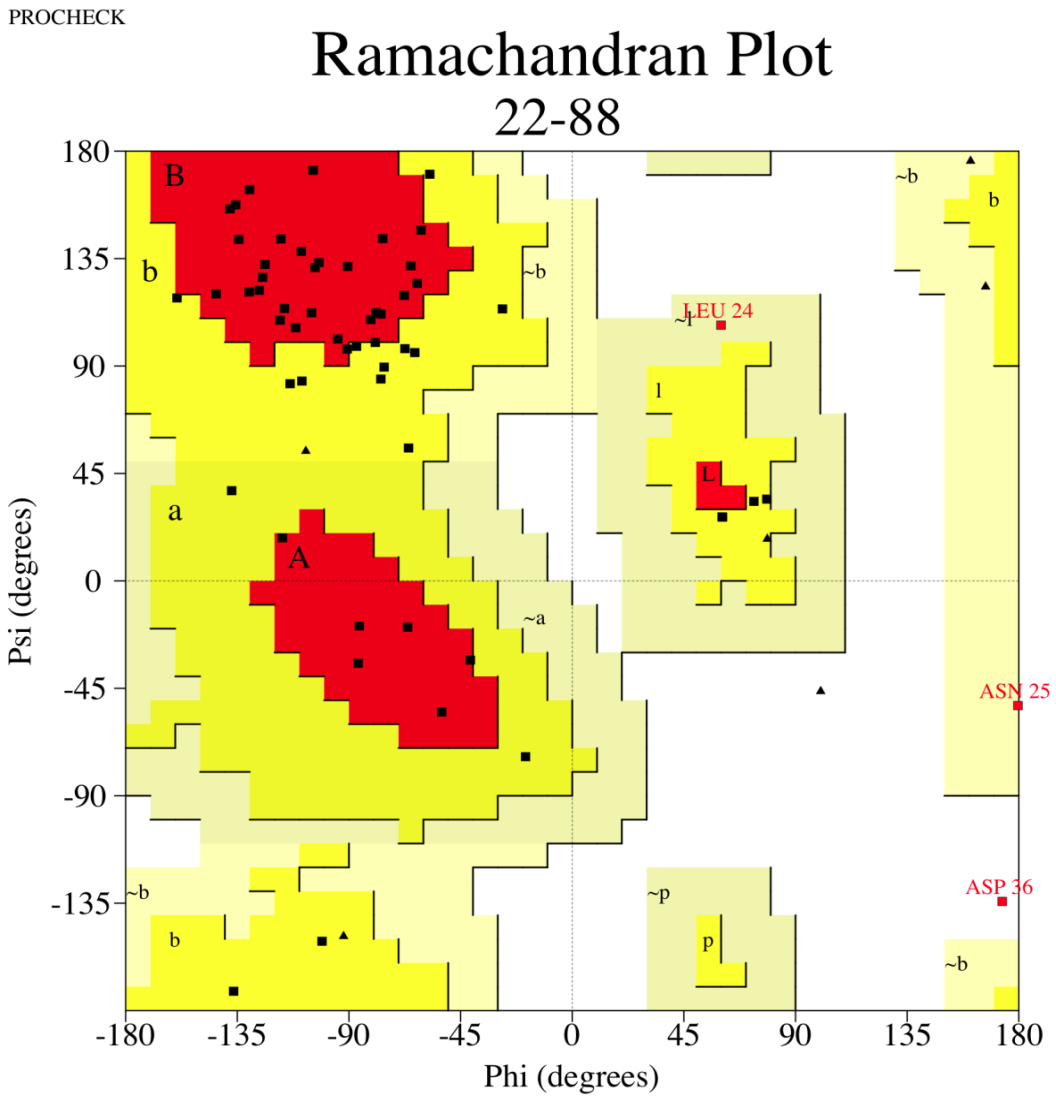
```
! resid 999 contains the four atoms which define the tensor
! the coordinates of resid 999 are appended to the pdb
! use our pdf file for the tensor
!
! the data are in the format:
! assign (atomselection) value error
      ( resid 12  and name HN  )   -1.86  0.500
      ( resid 13  and name HN  )    4.25  0.500
      ( resid 14  and name HN  )    12.62  0.500
      ( resid 15  and name HN  )   -3.02  0.500
      ( resid 16  and name HN  )   -2.36  0.500
      ( resid 22  and name HN  )   -7.21  0.500
      ( resid 26  and name HN  )   -3.4  0.500
      ( resid 27  and name HN  )    -3.42  0.500
      ( resid 28  and name HN  )    -0.69  0.500
      ( resid 29  and name HN  )    7.97  0.500
      ( resid 30  and name HN  )    -1.03  0.500
      ( resid 31  and name HN  )    -1.29  0.500
      ( resid 32  and name HN  )    -4.21  0.500
      ( resid 33  and name HN  )     4.25  0.500
      ( resid 35  and name HN  )   -7.1  0.500
      ( resid 36  and name HN  )   -0.34  0.500
      ( resid 38  and name HN  )   0.37  0.500
      ( resid 39  and name HN  )   -0.88  0.500
      ( resid 42  and name HN  ) 3.66  0.500
      ( resid 43  and name HN  )   -0.44  0.500
      ( resid 44  and name HN  )    7.29  0.500
      ( resid 45  and name HN  )    -2.57  0.500
      ( resid 46  and name HN  )    -9.19  0.500
      ( resid 47  and name HN  )    -5.28  0.500
      ( resid 48  and name HN  )   -3.99  0.500
      ( resid 49  and name HN  )    -13.55  0.500
      ( resid 50  and name HN  )    -1.81  0.500
      ( resid 51  and name HN  )    -1.68  0.500
      ( resid 53  and name HN  )    5.51  0.500
      ( resid 54  and name HN  )   -0.96  0.500
      ( resid 55  and name HN  )   -8.81  0.500
      ( resid 56  and name HN  )   6.03   0.500
      ( resid 57  and name HN  )   -1.47  0.500
      ( resid 58  and name HN  )   -3.03  0.500
      ( resid 61  and name HN  )   7.82  0.500
      ( resid 62  and name HN  )   -8.65  0.500
      ( resid 63  and name HN  )   -5.78  0.500
      ( resid 64  and name HN  )   -1.59  0.500
      ( resid 66  and name HN  )   -2.16  0.500
      ( resid 69  and name HN  )   15.01  0.500
      ( resid 70  and name HN  )   5.41  0.500
      ( resid 72  and name HN  )   1.98  0.500
      ( resid 74  and name HN  )   10.41  0.500
      ( resid 75  and name HN  )   -1.01  0.500
      ( resid 76  and name HN  )   6.34  0.500
      ( resid 77  and name HN  )   -5.94  0.500
      ( resid 78  and name HN  )   -3.34  0.500
      ( resid 79  and name HN  )   -1.97  0.500
      ( resid 80  and name HN  )   -6.4  0.500
      ( resid 81  and name HN  )   0.35  0.500
      ( resid 82  and name HN  )   -5.12  0.500
      ( resid 83  and name HN  )   0.43  0.500
      ( resid 84  and name HN  )   1.54  0.500
```

```
( resid 85  and name HN  )   11.97  0.500
( resid 86  and name HN  )   0.94  0.500
( resid 87  and name HN  )   0.74  0.500
( resid 88  and name HN  )   0.61  0.500
( resid 91  and name HN  )   -5.4  0.500
```

Appendix 5 Dihedral angle restraints for OutC-HRF3

13 A -58.509 -33.955 5.576 8.313 10.197 10 Good
14 S -63.786 -35.411 8.127 11.832 9.263 10 Good
15 Q -69.240 -34.819 8.748 13.788 8.082 10 Good
28 L -95.855 114.273 12.294 10.098 17.738 10 Good
31 V -135.764 135.817 14.646 15.265 26.342 10 Good
32 M -111.572 122.096 19.252 10.052 17.522 10 Good
33 A -88.860 132.354 28.151 16.517 19.236 10 Good
37 D -60.678 -34.532 3.960 8.178 19.693 10 Good
38 S -68.758 -21.450 9.195 17.939 13.606 10 Good
42 A -117.713 127.344 11.759 12.941 18.709 10 Good
43 I -101.268 127.373 14.046 8.984 19.332 10 Good
44 I -116.449 134.994 11.823 12.515 17.177 10 Good
45 S -112.405 123.524 12.521 7.200 15.218 10 Good
46 K -138.046 125.749 12.479 9.752 30.817 10 Good
48 N 70.351 10.815 10.372 13.646 37.868 10 Good
49 E -113.579 144.558 15.571 15.022 29.882 10 Good
50 Q -105.667 137.701 18.447 13.934 15.890 10 Good
51 F -132.786 146.504 21.942 16.036 13.969 10 Good
52 S -109.918 122.316 13.050 9.667 22.850 10 Good
53 R -123.582 155.819 9.168 9.456 31.207 10 Good
62 Y -123.379 152.031 11.947 13.574 35.852 10 Good
64 A -109.287 136.763 22.377 24.116 21.921 10 Good
65 K -125.234 136.228 20.830 14.924 19.047 10 Good
66 I -78.586 126.444 11.711 14.383 17.935 10 Good
68 S -154.836 149.824 8.922 10.961 20.075 10 Good
69 I -125.245 120.470 21.015 12.099 26.141 10 Good
70 R -96.782 127.157 22.680 28.512 61.187 10 Good
74 V -129.757 142.284 16.583 19.211 21.654 10 Good
75 V -114.601 123.436 11.309 10.422 15.597 10 Good
76 L -113.988 138.792 12.672 16.023 12.848 10 Good
77 Q -118.803 123.987 14.595 10.902 16.668 10 Good
79 Q 52.520 43.379 3.762 3.833 42.229 10 Good
80 G 73.402 9.413 9.489 12.054 36.762 10 Good
81 R -112.242 150.472 16.410 15.550 32.751 10 Good
82 Y -124.155 140.285 14.232 15.704 22.066 10 Good
83 E -138.683 158.208 14.629 11.432 16.346 10 Good
84 V -128.124 132.775 16.185 13.197 16.336 10 Good
85 L -116.602 133.278 15.302 15.774 24.463 10 Good
87 L -89.803 138.927 24.140 16.407 24.403 10 Good
88 Y -94.083 129.382 19.055 9.613 19.985 10 Good
89 S -99.219 138.600 28.685 12.731 16.789 10 Good
92 D -83.143 -20.845 22.359 23.961 9.961 10 Good
96 D -94.542 -12.761 21.488 24.294 22.030 10 Good
97 G 81.596 14.899 8.139 18.877 29.391 10 Good
98 V -90.361 120.495 24.230 27.602 68.050 10 Good
99 P -58.852 140.560 6.688 12.172 64.785 10 Good

Appendix 6 OutC-HRF3 best 20 structures PROCHECK Ramachamdran plot



Plot statistics

Residues in most favoured regions [A,B,L]	34	60.7%
Residues in additional allowed regions [a,b,l,p]	19	33.9%
Residues in generously allowed regions [~a,~b,~l,~p]	2	3.6%
Residues in disallowed regions	1	1.8%
Number of non-glycine and non-proline residues	56	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	6	
Number of proline residues	2	
Total number of residues	66	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.

Appendix 7 Paper to be published in *The Journal of Biological Chemistry*

Solution structure of the HR domain of the type II secretion system

Shuang Gu ‡, Geoff Kelly§, Xiaohui Wang¶, Tom Frenkiel§, Vladimir E. Shevchik¶1, and Richard W. Pickersgill‡

From the‡School of Biological and Chemical Sciences, Queen Mary University of London, London E1 4NS, United Kingdom, §MRC Biomedical NMR Centre, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, United Kingdom, and ¶The Universite´ de Lyon, F-69003, Universite´ Lyon 1, Lyon, F-69622, INSA-Lyon, Villeurbanne F-69621, CNRS, UMR5240, Microbiologie Adaptation et Pathoge´nie, Lyon, F-69622, France

THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 287, NO. 12, pp. 9072–9080, March 16, 2012

Solution Structure of Homology Region (HR) Domain of Type II Secretion System*[S]

Received for publication, September 2, 2011, and in revised form, January 13, 2012 Published, JBC Papers in Press, January 17, 2012, DOI 10.1074/jbc.M111.300624

Shuang Gu[‡], Geoff Kelly[§], Xiaohui Wang[¶], Tom Frenkiel[§], Vladimir E. Shevchik[¶1], and Richard W. Pickersgill[‡2]

From the [‡]School of Biological and Chemical Sciences, Queen Mary University of London, London E1 4NS, United Kingdom, [§]MRC Biomedical NMR Centre, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, United Kingdom, and the [¶]Université de Lyon, F-69003, Université Lyon 1, Lyon, F-69622, INSA-Lyon, Villeurbanne F-69621, CNRS, UMR5240, Microbiologie Adaptation et Pathogénie, Lyon, F-69622, France

Background: The HR domain is essential for the function of the T2SS.
Results: HR domain has similar structure to PilP of the T4PS and interacts with secretin.
Conclusion: The evolutionary link between T2SS and T4PS extends to proteins of the inner membrane subcomplex.
Significance: This work is crucial to understanding how this machine works and how it may be stopped.

The type II secretion system of Gram-negative bacteria is important for bacterial pathogenesis and survival; it is composed of 12 mostly multimeric core proteins, which build a sophisticated secretion machine spanning both bacterial membranes. OutC is the core component of the inner membrane subcomplex thought to be involved in both recognition of substrate and interaction with the outer membrane secretin OutD. Here, we report the solution structure of the HR domain of OutC and explore its interaction with the secretin. The HR domain adopts a β-sandwich-like fold consisting of two β-sheets each composed of three anti-parallel β-strands. This structure is strikingly similar to the periplasmic region of PilP, an inner membrane lipoprotein from the type IV pilus system highlighting the common evolutionary origin of these two systems and showing that all the core components of the type II secretion system have a structural or sequence ortholog within the type IV pili system. The HR domain is shown to interact with the N0 domain of the secretin. The importance of this interaction is explored in the context of the functional secretion system.

Gram-negative bacteria have evolved multiple sophisticated machines to transport selected proteins across the two bacterial membranes. The type II secretion system (T2SS)[3] is specifically used to secrete folded proteins such as lytic enzymes and toxins from the periplasm to the external medium or to host tissue (1, 2). This system is shared by many bacteria pathogenic for human, animals, and plants where it plays a pivotal role in pathogenesis. The T2S machinery is composed of 12 to 15 components with generic names GspA to GspO. The T2SS components of the phytopatogen bacteria *Dickeya dadantii* (formerly *Erwinia chrysanthemi*) are named Out. Most of these proteins are multimeric and build together a sophisticated machine spanning both bacterial membranes. The machine can be divided into three subassemblies. The secretin GspD (Gsp is the nomenclature for the enterotoxigenic *Escherichia coli* and Out for *D. dadantii* proteins) forms large gated channels in the outer membrane through which the proteins are secreted (3, 4). Five pseudopilins (GspG to GspK) are processed by the prepilin peptidase/methylase GspO and form a short pilus not exceeding the periplasm, which could push proteins to be secreted through the pore formed by the secretin. GspC, GspL, GspM, and GspF are thought to form a subcomplex in the inner membrane (5). An ATPase, GspE is associated with this subcomplex and provides energy to the secretion process probably by assuring assembly of the pseudopilus (6). However, the exact role of individual components and the molecular organization of the T2S machinery have remained elusive.

Phylogenetic analyses suggested that the T2SS is ancestrally related to the type IV pili system (T4PS) and share several homologous core elements (7). The current models of T2S also suggest its functional similarity with T4P system, notably, that short pseudopilus assembled in the periplasm push the protein across the pore formed by the secretin. Because the inner membrane components of T2SS and T4P, GspL/M/C, and PilM/N/O/P, respectively, do not share any sequence similarity, they have been assumed to execute functions specific for each of these systems (1, 8). However, recent structural analysis

* This work was supported by the Biology and Biotechnology Sciences Research Council (U.K.), Higher Education Funding Council for England, and Queen Mary University of London.
The atomic coordinates and structure factors (code 2LNV) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).
NMR spectra and assignments have been deposited in the BioMag ResBank under accession no. 18181.
[S] This article contains supplemental Tables S1 and S2 and Figs. S1–S4.
[1] Supported by a grant from French ANR-2010-BLANC-1531 SecPath program. To whom correspondence may be addressed: Université de Lyon, F-69003, Université Lyon 1, Lyon, F-69622, INSA-Lyon, Villeurbanne F-69621, CNRS, UMR5240, Microbiologie Adaptation et Pathogénie, Lyon F-69622, France. Tel.: 04-72-44-58-27; Fax: 04-72-43-15-84; E-mail: vladimir.shevchik@insa-lyon.fr.
[2] To whom correspondence may be addressed: School of Biological and Chemical Sciences, Queen Mary University of London, London E1 4NS, UK. Tel.: 44-0-20-7882-8444; Fax: 44-0-20-7882-7732; E-mail: r.w.pickersgill@qmul.ac.uk.

[3] The abbreviations used are: T2SS, type II secretion system; T4PS, type IV pili system; HR, homology region; r.m.s.d., root mean square deviation.

revealed that the periplasmic domains of GspL and PilN on one side and GspM and PilO on the other adopt the same ferredoxin-like fold and hence could play similar functions within T2S and T4P systems, respectively (9–11). Moreover, the cytoplasmic region of GspL adopts the actin-like fold of PilN (12). Therefore, GspC is yet the sole T2SS core component that has not a sequence or structural ortholog within the T4PS. GspC is a bitopic inner membrane protein. The single trans-membrane segment drives dimerization of the protein (13). The periplasmic region possesses a so-called homology region (HR) and, in most known GspCs, including OutC, a PDZ domain (see Fig. 1*A*). The PDZ domain is involved in secretion specificity but is not an essential module because in certain GspCs, it is absent or replaced by a coiled-coil domain (14). The exact function of HR is unknown; however, it was shown to interact *in vitro* with the N-terminal region of the secretin GspD and especially with the N0 domain (15, 16).

Secretins constitute the outer membrane pore necessary for translocation of macromolecules. Beyond the T2SS system, secretins are also involved in T4P, T3SS and filamentous phage release (17). Secretins consist of two main regions, a variable N-terminal periplasmic region and a conserved pore-forming C-domain, also named the secretin domain. The N-terminal region has a modular structure and in secretins from different systems possesses a variable number of N0, N1/N2, and N3 domains (18, 19). A schematic of the domain organization of OutD the cognate secretin for OutC is also shown in Fig. 1*A*. This N-terminal region is thought to assure specific functions within each of the export pathways. Notably, it interacts with the inner membrane components of secretion systems and with the proteins to be exported (4, 19, 20).

In this work, we report determination of the solution structure of the homology region domain of OutC and investigate its interaction with the secretin OutD from the plant pathogenic bacteria *Dickeya dadantii* both *in vitro* and *in vivo*. This study reveals a striking structural similarity between the HR domain and the periplasmic domain of PilP from the T4PS, revealing that these important components of the inner membrane subcomplexes evolved from a common ancestor; this strengthens the view that these molecular machines have a common evolutionary origin. We also show that β1 of the HR domain interacts with the N0 domain *in vitro* and explore the importance of the interactions involved *in vivo*.

EXPERIMENTAL PROCEDURES

Protein Expression, Purification, and Analysis—The HRF3 fragment was discovered using limited proteolysis of the protein produced using the pGX-oC-HR construct (supplemental Table S1). Sequencing grade trypsin was used. The protein sample was bound to a C5 reverse-phase HPLC column in the presence of 1% formic acid solution and eluted using an acetonitrile gradient. Three fractions from the HPLC column were characterized by mass spectroscopy, HRF1, -2, and -3; HRF3 was the shortest of these nested fragments and was therefore selected for further study. N0, N1, and N1-N2 domains of OutD were produced in *E. coli* and purified by nickel-affinity chromatography as described previously (13).

The HRF3 DNA was amplified from HR construct (pGX-oC-HR) and cloned into the pET14b expression vector (strains and primers used are detailed in supplemental Tables S1 and S2). The expression vector was transformed into competent BL21 (DE3) *E. coli* cells. Following bacterial growth and induction the protein was loaded onto a 5-ml His Trap HP column. After washing, Thrombin (GE Healthcare) was used to release the His tag cleaved HRF3 from the column. HRF3 was then passed through a Superdex75 column (GE Healthcare) column. The fragment eluted at a volume consistent with HRF3 being a subunit in solution. This was confirmed by dynamic light scattering. For NMR spectroscopy, uniformly ^{15}N- and ^{13}C-labeled proteins were produced by growing cell cultures in M9 minimal medium that contained 1 g/liter ^{15}N-ammonium chloride and 2 g/liter ^{13}C-D-glucose (Cambridge Isotope Laboratories, Inc.) as the sources of nitrogen and carbon, respectively.

NMR Spectroscopy—NMR spectra were acquired at 15 °C using Bruker Avance 700 and 600 MHz spectrometers equipped with CryoProbes. NMR samples had a typical concentration of 0.3–0.7 mM in 20 mM Tris, pH7.0, and 10% D_2O. 150 mM NaCl was included in the cross-titration studies. Fast chemical shift-based structure calculations were used initially employing CS-Rosetta (21, 22). All spectra were processed using NMRPipe/NMRDraw (23) and analyzed using XEASY (24). HNCA, HN(CO)CA, HNCO, HNCACB, and CBCA(CO)NH experiments were employed to obtain sequence-specific ^{1}HN, ^{15}N, ^{13}Cα, ^{13}Cβ, and ^{13}C′ backbone assignments. Side chain aliphatic proton and carbon assignments were achieved by a combination of three-dimensional ^{15}N-edited total correlation spectroscopy- and NOESY-heteronuclear single-quantum correlation spectroscopy (25). Secondary structure elements were determined from a combination of ^{1}HN, ^{15}N, ^{13}Cα, ^{13}Cβ, and ^{13}C′ secondary chemical shifts using TALOS+ (26, 27) as well as from identification of interstrand NOE contacts. Hydrogen bonds were measured by direct trans-hydrogen bond (N-H–O=C) scalar coupling (28). Structure calculation used ARIA (version 1.2) (29) and input NOE data from three-dimensional ^{15}N-edited NOESY-HSQC, ^{13}C-edited NOESY-HSQC (all acquired with a mixing time of 100 ms), 76 φ and ψ torsion angles were derived from TALOS and residual dipolar coupling measurements. The standard ARIA protocol was used with nine iterations and 100 structures generated per iteration. In the final iteration, 30 structures with the lowest energy were used for water refinement and 20 structures with the lowest energy were selected as representatives of OutC-HRF3 structure.

In Vivo Analysis of Mutant OutC and OutD Variants—To test the functionality of OutC and OutD variants carrying site-specific substitutions, complementation assays were used with *D. dadantii* A3556 Δ*outC* and A3558 Δ*outD* strains. Bacteria carrying an appropriate pTdB plasmid with a mutant allele of either *outC*, or *outD*, or *outC-outD* were grown in Luria-Bertani medium supplemented with sodium galacturonate at 1 g·liter^{-1} at 30 °C aerobically at 150 rpm to steady-state; then, culture supernatant and cells were separated by centrifugation and analyzed by immunoblotting with PelD and PelI antibodies as described previously (13). Abundance of the OutC and OutD mutant variants in the cells was assessed by immunoblotting

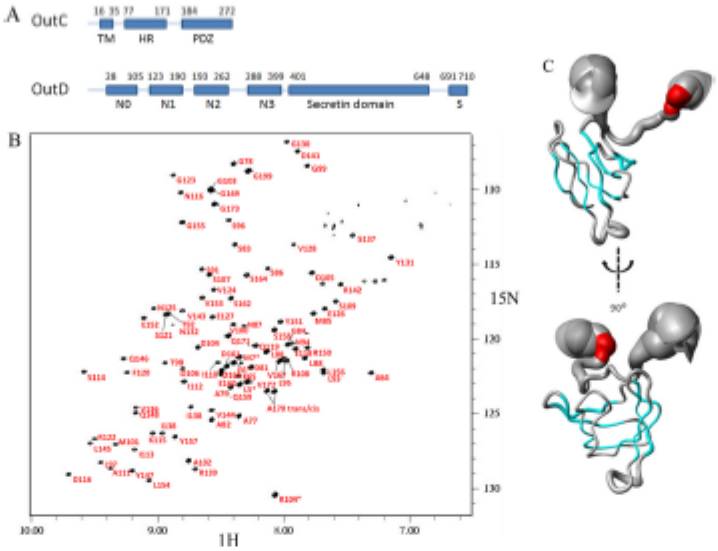


FIGURE 1. **Tertiary structure of the HR domain of OutC.** *A*, schematic representation of the domain organization of inner membrane subcomplex component OutC and secretin OutD. *B*, assignment of the ^{1}H-^{15}N HSQC spectra for the HR domain. *C*, sausage representation of the best 20 calculated OutC-HRF3 structures. The secondary structure is shown in color: α-helical residues are in *red*, and the β-sheet is in *cyan*. The diameter of the sausage reflects the dynamics of the protein in solution, which is plotted according to the residue α-carbon T2 relaxation profile.

with OutC and OutD antibodies. For *in vivo* disulfide bonding analysis, *D. dadantii* cells carrying appropriate plasmids with mutant *outC* and/or *outD* genes were grown as above. Cells were washed in TBS (50 mM Tris-HCl, pH 7.5, 100 mM NaCl), and remaining free thiol groups were blocked with 20 mM iodoacetamide in TBS at 25 °C for 30 min. The cells were then pelleted, resuspended in Laemmli sample buffer without 2-mercaptoethanol and boiled for 10 min, and the extent of disulfide bonding was assessed by SDS-PAGE, followed by immunoblotting with anti-OutC and anti-OutD antibodies.

RESULTS

Identification of Folded Homology Region Domain (OutC-HR)—OutC comprises a short cytoplasmic sequence, a transmembrane helix, and two periplasmic regions: the homology region and a PDZ domain (Fig. 1*A*). The HR region is so named because a similar sequence is found in all GspC proteins. Examination of the ^{1}H-^{15}N HSQC spectrum of the periplasmic region of OutC (residues 60 to 272) reveals that in addition to the PDZ domain, there is also another folded domain. Limited proteolysis of this OutC derivative gave three bands (HRF1, HRF2, and HRF3) on both SDS-PAGE and native gel (supplemental Fig. S1). The native gel was used to assess the solubility of the proteolytic fragments as insoluble protein tends to aggregate and not enter the gel. The precise identities of the fragments were established by mass spectroscopy (supplemental Fig. S2) and confirmed by N-terminal sequencing. The smallest fragment (HRF3), corresponding to residues 77 to 172, was produced recombinantly by sub-cloning the appropriate part of the *outC* gene. The HRF3 protein produced in this way was both soluble and stable and was used for these studies. The ^{1}H-^{15}N HSQC spectrum of HRF3 was well dispersed and overlaid well with the previously measured spectrum for the entire periplasmic region of OutC, confirming that HRF3 comprises an independently folded domain (supplemental Fig. S3).

Solution Structure of Homology Region Domain of OutC—NMR relaxation measurements revealed that the structured region of HRF3 comprised 67 residues (residues 91 to 157), and we refer to this domain as the homology region (HR) domain. A nearly complete sequence-specific NMR signal assignment was achieved using a combination of three-dimensional triple resonance experiments using ^{15}N,^{13}C-labeled OutC-HR, and three-dimensional NOESY-HSQC, TOCSY-HSQC using ^{15}N-labeled protein (Fig. 1*B*). The 1954 experimental NOE distance restraints were obtained using isotope-edited NOESY spectra, and 76 φ and ψ torsion angle restraints were derived from secondary chemical shift analysis using the TALOS algorithm. Twenty eight hydrogen bonds were measured experimentally. The structure of the folded domain was determined with high precision as judged from the overall root mean square deviation (r.m.s.d.) of 0.46 and 1.26 Å for the backbone and heavy atoms, respectively (Table 1) for the pairwise r.m.s.d. for the family of 20 representative structures. It is significant that even the loop regions connecting the anti-parallel β-strands are well determined in this structure (Fig. 1*C*). It is interesting to note that the structure predicted by CS-ROSETTA, using only the chemical shift data, was very close to that determined using conventional methods.

The architecture of the central 67 residues is β-sandwich-like comprising two three-stranded (up-down-up) anti-parallel β-sheets with short hairpin loops connecting adjacent β-strands within the two sheets and a longer loop between the two sheets (Fig. 2, *A–C*). The angle between the strands

TABLE 1
NMR refinement statistics for the HR domain (OutC-HRF3)

NMR distance and dihedral constraints	
Distance constraints	
Total NOE	1954
Intra-residue	652
Inter-residue	
Sequential ($\|i-j\|=1$)	381
Medium range ($\|i-j\|\geq 2\leq 4$)	170
Long range ($\|i-j\|\geq 5$)	751
Hydrogen bonds (experimental measured)	28
Hydrogen bonds(observed in >50% of structures)	64
Total dihedral angle restraints (TALOS)	76
φ	38
ψ	38
Structure statistics	
Violations (mean ± S.D.)	
Distance constraints (>0.5) (Å)	0
Dihedral angle constraints (>5) (°)	0
Deviations from idealized geometry	
Bond lengths (Å)	0.005 ± 0.0001
Bond angles (°)	0.671 ± 0.011
Impropers (°)	0.748 ± 0.024
Average pairwise r.m.s.d. (Å)[a]	
Backbone	0.46 ± 0.11
Heavy	1.26 ± 0.20
R.m.s.d. from the mean structure (Å)[a]	0.33 ± 0.07
Residues in allowed regions[b]	98.3%
Residues in disallowed regions	1.7%

[a] Pairwise r.m.s.d. and r.m.s.d. from the mean structure were calculated among the 20 lowest energy structures from 50 refined structures (amino acid residues 91 to 157).
[b] Structure quality was analyzed with MolProbability over structured regions (amino acid residues 91–157). One residue in a tight β-turn between β5 and β6 (Pro^{140}) is in a disallowed region.

in the two β-sheets is ~45°. In several members of the family of structures, there is a fourth β-strand in the second β-sheet. This strand is more dynamic in solution than the other six β-strands, which are present in each of the 20 structures. The first β-strand is involved in interaction with the N0 domain of OutD and is illustrated in Fig. 2*D* (see also results below). Aliphatic residues, mostly leucines, from the internal surface of the β-sheets form the hydrophobic core of the HR domain (Fig. 2*E*). To both the N- and C- terminal ends of the β-sandwich, there are ~20 residues that are less regular in structure.

HR Domain Is Structurally Similar to PilP—A Dali database search revealed that the HR domain is similar to the periplasmic domain of PilP, an inner membrane core component of the T4PS of *Neisseria meningitidis* (Protein Data Bank code 2IVW) (30). The Z score is 3.7, and root mean square deviation is 6.8 Å for 64 equivalent α-carbon atoms. The HR domain has the same β-strand topology and common β-sandwich type fold as PilP (Fig. 2, *A–C*). In some of the HR family of structures, the conserved β4 strand of PilP is frequently not formed revealing greater structural flexibility in this part of the HR domain. To aid comparison with PilP, and in light of this flexible "β4" region, we use the same β-strand numbering for HR as used previously for PilP. Sequence alignment of PilP homologs from various T4PS indicates that strand β4 together with short neighboring loops L4 and L5 constitute a more conserved

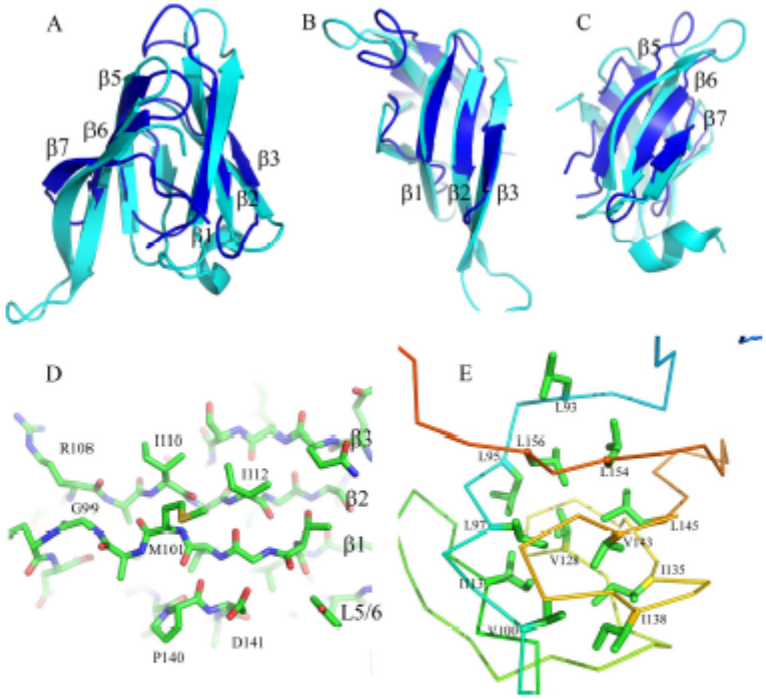


FIGURE 2. HR domain structure compared with PilP. *A*, schematic representation of the structure of the HR domain (*blue*) superimposed on PilP (*cyan*). The β-strands are labeled according to PilP numbering. *B* and *C* look down onto the first and second sheet of the HR domain, respectively. *D*, the first β-sheet in stick representation, highlighting β1, which is important in the interaction of HR with OutD-N0 (see Fig. 5). *E*, hydrophobic residues forming the core of the HR domain in stick representation with backbone displayed as a ribbon.

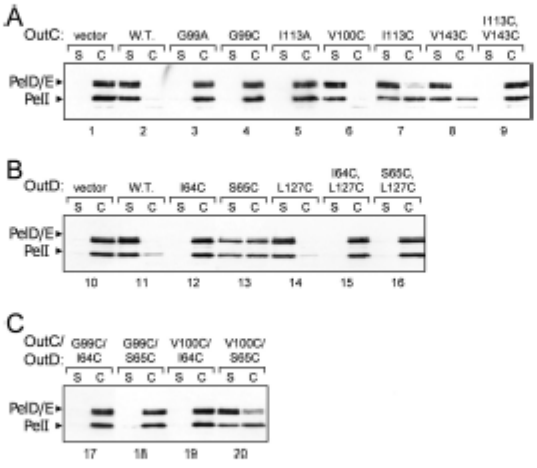


FIGURE 3. Functional analysis of OutC and OutD point mutants. The secretion activity of OutC and OutD mutants (indicated on *top* of gels) was estimated in a complementation assays with *D. dadantii* strains A3656 Δ*outC* (*A* and *C*) and A3658 Δ*outD* (*B*). Bacteria carrying a plasmid with mutant alleles of either *outC* (*A*), or *outD* (*B*), or *outC* and *outD* (*C*) were grown to steady-state; then, culture supernatant (*S*) and cells (*C*) were separated and analyzed by immunoblotting with PelD and PelI antibodies. The amount of secreted proteins (PelD, PelE, and PelI) in culture supernatant reflects secretion efficiency. The abundance of the corresponding OutC and OutD variants is shown on Fig. 4.

region than the corresponding region of the HR sequence (supplemental Fig. S4).

Several Conserved Residues Form Hydrophobic Core of HR—The two β-sheets composing the HR domain seem largely independent and are connected together through several hydrophobic interactions, including the side chains of well conserved Leu[93], Leu[97], Ile[113], Val[143], Leu[154], and Leu[156] and semiconserved Leu/Ile[145], Ile/Leu/Val[135], and Ile/Val[138] (Fig. 2*E* and supplemental Fig. S4). Most of these residues are also conserved in the PilP periplasmic domain highlighting their structural importance. Ile[113] and Val[143] contribute to the hydrophobic packing between the two sheets. The proximity of Ile[113] and Val[143] was assessed *in vivo* by site-directed mutagenesis and cysteine mediated cross-linking analysis. Substitution of one of these residues with cysteine did not affect notably the secretion in *D. dadantii*, whereas substitution with alanine impaired it (Fig. 3*A*, *lanes 5*, *7*, and *8*), indicating that in contrast to cysteine, alanine is not able to replace these hydrophobic side chains. A non-reducing gel revealed that OutC^{I113C} and OutC^{V143C} remained mostly as a non-cross-linked subunit (Fig. 4*A*, *lanes 4* and *5*), implying that the corresponding side chains are poorly accessible for intermolecular disulfide formation. Double I113C/V143C substitution fully prevented secretion (Fig. 3, *lane 9*) and provoked an efficient intramolecular disulfide bond within OutC (Fig. 4*A*, *lane 6*). Indeed, on non-reducing gel, OutC$^{I113C/V143C}$ migrated slightly faster indicative of its more compact shape. These mutational analyses are consistent with the solution structure determined for the HR domain.

OutC-HR Domain Interacts with OutD-N0 Domain—Previous NMR studies had demonstrated that the periplasmic part of OutC (residues 60 to 172 corresponding to the periplasmic region of OutC with the PDZ domain deleted) interacted with the N-terminal domains (N0, N1, N2, N3) of the secretin OutD (16). When unlabeled N0 domain (residues 28 to 112) was titrated into ^{15}N-labeled HR domain (residues 77 to 172), the peaks loosing intensity on the ^{1}H-^{15}N-HSQC spectra were those seen to lose intensity when the N-terminal domains of OutD were titrated into OutC$^{60-272}$ (supplemental Fig. S3). This confirms that it is the HR domain that is interacting with the periplasmic domains of OutD. However, shifts were not observed when the domain pair N1-N2 (residues 116 to 285) was titrated into the labeled HR domain. These data reveal that the N0 domain and not the N1-N2 domains interact with the HR domain. The major peak shifts on the ^{1}H-^{15}N-HSQC spectra when N0 was added involved β1 of the HR domain (Fig. 5*A*).

Non-linear sampling methods were then used to collect data from ^{13}C,^{15}N-labeled N0 domain allowing the assignment of 65 of the 80 residues. The backbone chemical shifts confirmed that the secondary structure of *D. dadantii* N0 was the same as that of *E. coli* N0, and a reasonable model of the *D. dadantii* N0 domain could therefore be built using the *E. coli* domain as a template. The chemical shifts observed when unlabeled HR domain was added could now be mapped to the surface of the N0 domain. The chemical shifts observed reveal that residues on strands β1, β2, and β3 of the HR domain interact with N0 with residues on β1 most affected (Fig. 5). Residues on β1 and α2 of the N0 domain were the most affected, but the shifts were relatively small. Small shifts were seen for residues: Thr[63], Ile[64], Ser[65], Phe[79], Ser[83], and Val[84]. A model of the HR·N0 complex was produced using HADDOCK (31) and is shown in Fig. 5*B*. The model brings β1 of HR and β3 of N0 together such that they could form a continuous anti-parallel β-sheet across both domains HR and N0. It is important to note that this is only a model of the interaction and the position and conformation of interfacial residues has not been defined experimentally. Met[101] appears to be an important hydrophobic residue at the interface with the more remote Asp[141] also showing large chemical shifts. The interaction between the *D. dadantii* HR and N0 domains is weak, but it does persist in the presence of 150 mM sodium chloride. The biological relevance of the interaction is assessed in the next section.

Assessment of HR Residues Involved in HR-N0 Interaction in Vivo—NMR cross-titration studies indicated that the HR-N0 interface includes the first β-strand of HR, notably Gly[99], Val[100], and Met[101] (Figs. 2*D* and 5). Moreover, the latter residue and Asp[141] showed large chemical shifts on binding. To explore the functional importance of these residues, they were substituted with alanine and cysteine and assessed in secretion assays (Fig. 3*A*). The D141A substitution was fully functional, indicating that this conserved residue (supplemental Fig. S4) is not crucial for the interaction with N0 and therefore has some other functional significance. Similarly, substitution of Val[100] with cysteine did not affect the secretion (Fig. 3*A*, *lane 6*), implying that cysteine can be accommodated in place of the valine without perturbation of the β-sandwich architecture. In contrast, Gly[99] and Met[101] could not be substituted without total loss of secretion function (Fig. 3*A*), indicating their crucial structural and functional roles, respectively. Substitution of Met[101] with alanine would lead to a loss of hydrophobic surface affecting

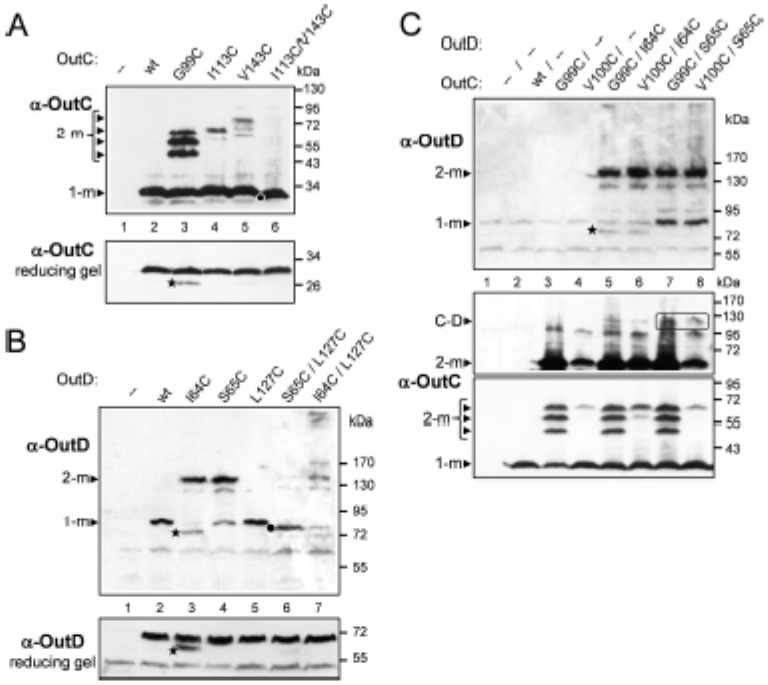


FIGURE 4. Disulfide cross-linking analysis of OutC and OutD interactions within the functional T2SS. *A*, disulfide cross-linking analysis of the HR domain of OutC. The G99C substitution provokes efficient homodimerization of OutC, whereas I113C or V143C single substitutions do not. Double substitution I113C/V143C leads to an intramolecular disulfide within OutC. *B*, disulfide cross-linking analysis of N0-N1 interaction sites in OutD. I64C and S65C single substitutions provoke efficient homodimerization of OutD, whereas L127C substitution does not. Double substitution S65C/L127C causes an intramolecular disulfide between N0 and N1 domains of OutD, whereas I64C/L127C does not. *C*, disulfide bonding analysis of HR-N0 interacting site. Combination of G99C and V100C substitutions in OutC with I64C and S65C substitutions in OutD does not provoke an efficient cross-linking between OutC and OutD. Some amounts of putative OutC-OutD complex are surrounded. *D. dadantii* strains A3556 Δ*outC* (*A* and *C*) and A3558 Δ*outD* (*B*) carrying a plasmid with mutant alleles of either *outC* (*A*), or *outD* (*B*), or *outC* and *outD* (*C*) (cysteine substitutions indicated on *top*) were grown aerobically to steady-state. Remaining free thiol groups were blocked with iodoacetamide, and the extent of disulfide bonding was assessed by SDS-PAGE under non-reducing conditions followed by immunoblotting with anti-OutC and anti-OutD antibodies. The same samples but treated with 2-mercaptoethanol were used to estimate abundance of the corresponding OutC and OutD variants (reducing gel). In *C*, the *middle panel* was overexposed in comparison with the *lower panel* to show high molecular mass species. Positions of monomers (*1-m*) and dimers (*2-m*) and putative OutC-OutD complex (*C* and *D*) are indicated by *arrowheads*. *Stars* indicate degradative products of OutCG99C and OutDI64C variants. *Dots* indicate OutCI113C/V143C and OutDS65C/L127C variants that migrate slightly faster than the corresponding wild-type monomers, indicating formation of an intramolecular disulfide bond.

interactions involving the HR-β1 interface. Gly^{99} is the most conserved residue among HR and PilP (supplemental Fig. S4). It is located in the center of β1 strand and introduces a bend to facilitate formation of β-bulge. Contrary to the other single cysteine substitutions in HR, $OutC^{G99C}$ formed dimers very efficiently (Fig. 4*A*, compare *lane 3* with *lanes 4* and *5*, Fig. 4*C*, *lower panel*, compare *lane 3* with *4*), suggesting that in the assembled secretion system, neighboring HR domains might interact via the first β-strand.

In Vivo Assessment of HR-N0 and N0-N1 Interactions—To probe *in vivo* proximity of β1 of HR with β3 of N0 domain, several residues located on these two β-strands were simultaneously substituted with cysteine in an attempt to isolate a disulfide bridged OutC and OutD pair. Single cysteine substitution of Ile^{64} and Ser^{65} on β3 of N0 provoked homodimerization of the mutant OutD (Fig. 4*B*, *lanes 3* and *4*), suggesting that in the functional T2SS, neighboring N0 domains could interact together through the β3 strand. When these mutations in OutD were combined pairwise with G99C and V100C substitutions in OutC, only low amounts of OutC-OutD complex was observed with $OutC^{G99C}$/$OutD^{S65C}$ couple (Fig. 4*C*, *middle panel*, *lane 7*). Perhaps an optimal arrangement of residues or the distance adequate for disulfide formation was not achieved.

We further probed by cysteine cross-linking analysis the crystal interface observed previously between the β3 of N0 and β6 of N1. This interface includes several pairs of proximal residues, notably Ser^{65} and Leu^{127}, on β3 and β6, respectively. Single L127C substitution behaved as the wild-type protein, was fully functional and was monomeric (Fig. 3*B*, *lane 14*, and Fig. 4*B*, *lane 5*). However, double S65C/L127C substitution became non functional (Fig. 3*B*, lane 16) and formed an intramolecular disulfide bond because it migrated faster than monomeric OutD (Fig. 4*B*, compare *lanes 5* and *6*). In contrast, L127C combined with I64C did not generate such an intramolecular cysteine bonding (Fig. 4*B*, *lane 7*). These data are compatible with mutual arrangement of the assessed residues within the N0/N1 crystal interface, where Ile^{64} is buried, whereas Ser^{65} and Leu^{127} are exposed to the interface. These

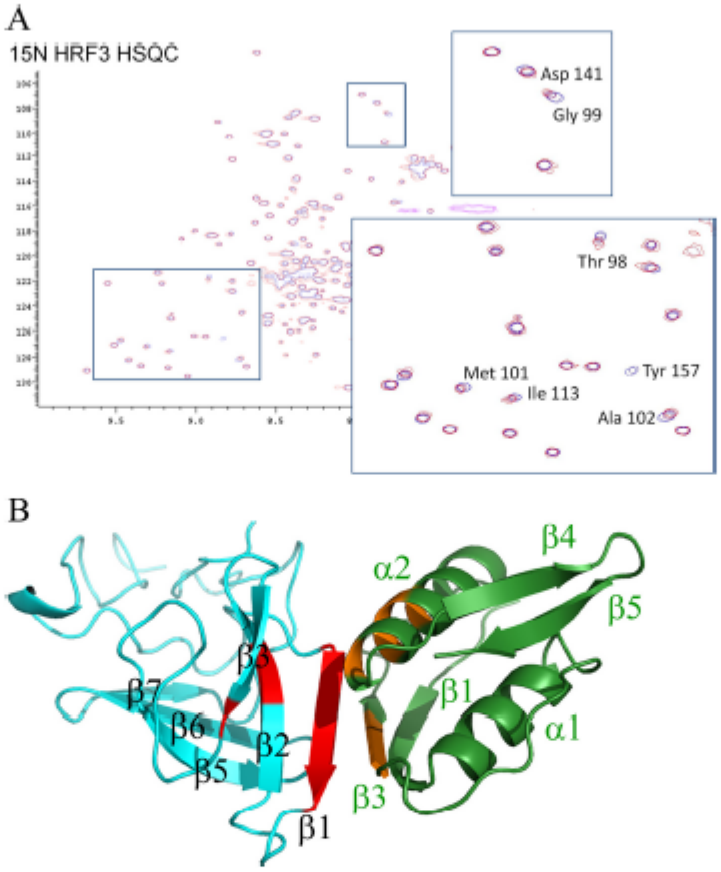


FIGURE 5. HADDOCK model of the OutC-HR·OutD-N0 complex in schematic representation. *A*, chemical shifts measured for the ^{15}N-HR domain when unlabeled N0 is added. *B*, the HR domain is on the *left* in *cyan*, and the N0 domain on the *right* in *green*. Chemical shifts are mapped onto the structures of HR domain in *red* and onto the N0 domain in *orange*, respectively. β1 of the HR domain interacts with β3 and α2 of N0. The peaks shifted on HRF3 are Thr^{98}, Gly^{99}, Val^{200}, Met^{101}, Ala^{102}, Ile^{112}, Ile^{113}, Phe^{120}, Asp^{141}, and Tyr^{88}. On N0, the peaks shifted are Thr^{62}, Ile^{64}, Ser^{65}, Phe^{79}, Ser^{82}, and Val^{84}. This is a modeled structure not an experimentally derived structure.

results reveal that β3 of the N0 domain can interact as well with an adjacent N0 domain (via β3-β3 strand interaction) as with N1 domain (via β3-β6 strand interaction) (Fig. 4*B*, *lanes 4* and *6*), suggesting a dynamic arrangement of the secretin N-terminal domains in the functional secretion system.

DISCUSSION

The similarity in structure of the HR domain of the T2SS and PilP of the T4PS suggests their evolution from a common ancestral domain, which strengthens the evolutionary link between these export machines, a link that did not previously extend to these core proteins of the inner membrane subcomplex. As both domains are located in the periplasm and are associated with machines that assemble pilins either to form a pilus (T4PS) or to push proteins through the outer membrane secretin (T2SS), the structural similarity may extend to a similarity in function. The HR domain binds to the N0 domain of the T2SS secretin and so by analogy, PilP has been shown to interact with the T4PS secretin stabilizing the complex (32).

It is clear from the *in vivo* studies that the interactions within the secretion system are dynamic. Probing the interaction between N0 and N1 of OutD in the functional secretion system confirmed that the interaction surface seen in the crystal complex of the N-terminal periplasmic domains of GspD with nanobody (18), mainly involving strands β3 and β6, corresponds to one of the functional arrangements of these domains *in vivo*, but we also discovered that the N0 domain can interact with a neighboring N0 domain through the same β3 strand. Interestingly, the crystal structure of the N-terminal region of EscC, the secretin from the type III secretion system, showed another mode of mutual arrangement of N0 and N1 domains (19), suggesting the dynamic character of interactions involving N-terminal domains of secretin. The binding of the OutC-HR domain to OutD-N0 must also be dynamic as β3 strand of N0 is

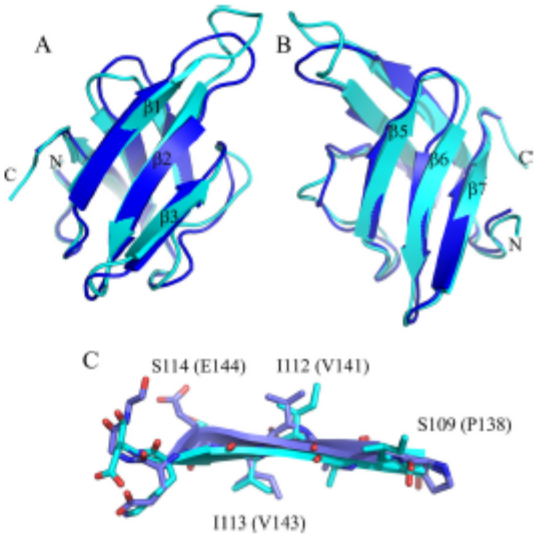


FIGURE 6. **Comparison of the solution structure and the crystal structure of the HR domain.** The solution structure of OutC-HR and crystal structure of GspD-HR are in *cyan* and *blue*, respectively. *A* and *B* are schematic representations looking down on to the first and second β-sheets. *C* illustrates the structural similarity between the structures for a central sheet strand, β2. Residue numbers are for OutC-HR, with those for GspD-HR in parentheses.

also detected in this interaction. When this paper was being published, the crystal structure of the enterotoxigenic *E. coli* GspC-HR-GspD-N0-N1 complex (Protein Data Bank code 3OSS) was described (33). The GspC-HR crystal structure is closely similar to our solution structure of OutC-HR (Fig. 6). The two HR domains have 35% sequence identity, and the 65 equivalent α-carbon atoms superimpose with an r.m.s.d. of 1.3 Å. In the crystal structure β1 of HR interacts with β1 of N0 and not β3 of N0. One explanation for this might be that the β3 region of N0 is particularly sticky and that it prefers to bind N1 but that in the absence of N1, it binds β1 of HR. There is clearly an argument that with both N0 and N1 domains present in the crystal structure, it will be more representative of the situation in the assembled secretion system, but both the solution structure and the crystal structure may be poor mimics of the organization present in the functional secretion system where the dodecameric arrangement of GspD/OutD subunits interacts with a likely hexameric or dodecameric arrangement of GspC/OutC subunits as well as with the other components of the T2SS and secreted proteins.

Our NMR study was initiated because the binding between the soluble regions of OutC and OutD was weak and so co-crystallization was not a possibility. The effectiveness of pulldown assays of OutC with OutD can be improved by co-expression or by partial unfolding in urea followed by refolding, but even then the complex is not stable in size exclusion chromatography even in conditions of low salt (16). The corresponding interaction for the enterotoxigenic *E. coli* GspC/GspD is clearly of higher affinity and can apparently be trapped more easily in the crystal. This apparently points to subtle differences between the secretion systems. Further research is clearly needed to establish the interactions between these and other domains in the context of the assembled secretion system and the way in which these interactions change during the secretion cycle.

In conclusion, this study completes the picture revealing that all core components of the T2SS have a structural ortholog within the T4PS. It also suggests the importance of β1 of the HR domain of the inner membrane subcomplex in interaction with β3 of the N0 domain of the secretin.

REFERENCES

1. Filloux, A. (2004) The underlying mechanisms of type II protein secretion. *Biochim. Biophys. Acta* 1694, 163–179
2. Cianciotto, N. P. (2005) Type II secretion: A protein secretion system for all seasons. *Trends Microbiol.* 13, 581–588
3. Chami, M., Guilvout, I., Gregorini, M., Rémigy, H. W., Müller, S. A., Valerio, M., Engel, A., Pugsley, A. P., and Bayan, N. (2005) Structural insights into the secretin PulD and its trypsin-resistant core. *J. Biol. Chem.* 280, 37732–37741
4. Reichow, S. L., Korotkov, K. V., Hol, W. G., and Gonen, T. (2010) Structure of the cholera toxin secretion channel in its closed state. *Nat. Struct. Mol. Biol.* 17, 1226–1232
5. Py, B., Loiseau, L., and Barras, F. (2001) An inner membrane platform in the type II secretion machinery of Gram-negative bacteria. *EMBO Rep.* 2, 244–248
6. Camberg, J. L., Johnson, T. L., Patrick, M., Abendroth, J., Hol, W. G., and Sandkvist, M. (2007) Synergistic stimulation of EpsE ATP hydrolysis by EpsL and acidic phospholipids. *EMBO J.* 26, 19–27
7. Peabody, C. R., Chung, Y. J., Yen, M. R., Vidal-Ingigliardi, D., Pugsley, A. P., and Saier, M. H. (2003) Type II protein secretion and its relationship to bacterial type IV pili and archaeal flagella. *Microbiology* 149, 3051–3072
8. Hazes, B., and Frost, L. (2008) Towards a systems biology approach to study type II/IV secretion systems. *Biochim. Biophys. Acta* 1778, 1839–1850
9. Sampaleanu, L. M., Bonanno, J. B., Ayers, M., Koo, J., Tammam, S., Burley, S. K., Almo, S. C., Burrows, L. L., and Howell, P. L. (2009) Periplasmic domains of *Pseudomonas aeruginosa* PilN and PilO form a stable heterodimeric complex. *J. Mol. Biol.* 394, 143–159
10. Abendroth, J., Rice, A. E., McLuskey, K., Bagdasarian, M., and Hol, W. G. (2004) The crystal structure of the periplasmic domain of the type II secretion system protein EpsM from *Vibrio cholerae*: The simplest version of the ferredoxin fold. *J. Mol. Biol.* 338, 585–596
11. Abendroth, J., Kreger, A. C., and Hol, W. G. (2009) The dimer formed by the periplasmic domain of EpsL from the type 2 secretion system of *Vibrio parahaemolyticus*. *J. Struct. Biol.* 168, 313–322
12. Karuppiah, V., and Derrick, J. P. (2011) Structure of the PilM-PilN inner membrane type IV pilus biogenesis complex from *Thermus thermophilus*. *J. Biol. Chem.* 286, 24434–24442
13. Login, F. H., and Shevchik, V. E. (2006) The single transmembrane segment drives self-assembly of OutC and the formation of a functional type II secretion system in *Erwinia chrysanthemi*. *J. Biol. Chem.* 281, 33152–33162
14. Bouley, J., Condemine, G., and Shevchik, V. E. (2001) The PDZ domain of OutC and the N-terminal region of OutD determine the secretion specificity of the type II out pathway of *Erwinia chrysanthemi*. *J. Mol. Biol.* 308, 205–219
15. Korotkov, K. V., Krumm, B., Bagdasarian, M., and Hol, W. G. (2006) Structural and functional studies of EpsC, a crucial component of the type 2 secretion system from *Vibrio cholerae*. *J. Mol. Biol.* 363, 311–321
16. Login, F. H., Fries, M., Wang, X., Pickersgill, R. W., and Shevchik, V. E. (2010) A 20-residue peptide of the inner membrane protein OutC mediates interaction with two distinct sites of the outer membrane secretin OutD and is essential for the functional type II secretion system in *Erwinia chrysanthemi*. *Mol. Microbiol.* 76, 944–955
17. Korotkov, K. V., Gonen, T., and Hol, W. G. (2011) Secretins: Dynamic channels for protein transport across membranes. *Trends Biochem. Sci.* 36, 433–443

18. Korotkov, K. V., Pardon, E., Steyaert, J., and Hol, W. G. (2009) Crystal structure of the N-terminal domain of the secretin GspD from ETEC determined with the assistance of a nanobody. *Structure* 17, 255–265
19. Spreter, T., Yip, C. K., Sanowar, S., André, I., Kimbrough, T. G., Vuckovic, M., Pfuetzner, R. A., Deng, W., Yu, A. C., Finlay, B. B., Baker, D., Miller, S. I., and Strynadka, N. C. (2009) A conserved structural motif mediates formation of the periplasmic rings in the type III secretion system. *Nat. Struct. Mol. Biol.* 16, 468–476
20. Shevchik, V. E., Robert-Baudouy, J., and Condemine, G. (1997) Specific interaction between OutD, an *Erwinia chrysanthemi* outer membrane protein of the general secretory pathway, and secreted proteins. *EMBO J.* 16, 3007–3016
21. Shen, Y., Lange, O., Delaglio, F., Rossi, P., Aramini, J. M., Liu, G., Eletsky, A., Wu, Y., Singarapu, K. K., Lemak, A., Ignatchenko, A., Arrowsmith, C. H., Szyperski, T., Montelione, G. T., Baker, D., and Bax, A. (2008) Consistent blind protein structure generation from NMR chemical shift data. *Proc. Natl. Acad. Sci. U.S.A.* 105, 4685–4690
22. Shen, Y., Vernon, R., Baker, D., and Bax, A. (2009) *De novo* protein structure generation from incomplete chemical shift assignments. *J. Biomol. NMR* 43, 63–78
23. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* 6, 277–293
24. Bartels, C., Xia, T. H., Billeter, M., Guntert, P., and Wuthrich, K. (1995). The program XEASY for computer-supported NMR spectral-analysis of biological macromolecules. *J. Biomol. NMR* 6, 1–10
25. Fesik, S. W., and Zuiderweg, E. R. (1988) Heteronuclear 3-dimensional NMR spectroscopy–A strategy for the simplification of homonuclear two-dimensional NMR-spectra. *J. Magn. Reson.* 78, 588–593
26. Shen, Y., and Bax, A. (2007) Protein backbone chemical shifts predicted from searching a database for torsion angle and sequence homology. *J. Biomol. NMR* 38, 289–302
27. Shen, Y., Delaglio, F., Cornilescu, G., and Bax, A. (2009) TALOS+: A hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR* 44, 213–223
28. Cordier, F., and Grzesiek, S. (1999) Direct observation of hydrogen bonds in proteins by interresidue (3h)(NC') scalar couplings. *J. Am. Chem. Soc.* 121, 1601–1602
29. Linge, J. P., Habeck, M., Rieping, W., and Nilges, M. (2003) ARIA: Automated NOE assignment and NMR structure calculation. *Bioinformatics* 19, 315–316
30. Golovanov, A. P., Balasingham, S., Tzitzilonis, C., Goult, B. T., Lian, L. Y., Homberset, H., Tønjum, T., and Derrick, J. P. (2006) The solution structure of a domain from the *Neisseria meningitidis* lipoprotein PilP reveals a new β-sandwich fold. *J. Mol. Biol.* 364, 186–195
31. Dominguez, C., Boelens, R., and Bonvin, A. (2003) HADDOCK: A protein-protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.* 125, 1731–1737
32. Balasingham, S. V., Collins, R. F., Assalkhou, R., Homberset, H., Frye, S. A., Derrick, J. P., and Tønjum, T. (2007) Interactions between the lipoprotein PilP and the secretin PilQ in *Neisseria meningitidis*. *J. Bacteriol.* 189, 5716–5727
33. Korotkov, K. V., Johnson, T. L., Jobling, M. G., Pruneda, J., Pardon, E., Heroux, A., Turley, S., Steyaert, J., Holmes, R. K., Sandkvist, M., and Hol, W. G. (2011) Structural and functional studies of the interaction of GspC and GspD in the type II secretion system. *PloS Pathog.* 7, e1002228

Appendix 8 Paper to published in PLoS Pathogens

Structural and Functional Insights into the Pilotin-Secretin Complex of the Type II Secretion System

Shuang Gu[1], Saima Rehman[1], Xiaohui Wang[2], Vladimir E. Shevchik[2*], Richard W. Pickersgill[1]

1 Queen Mary University of London, School of Biological and Chemical Sciences, London, England, **2** Université de Lyon, Université Lyon 1, Lyon; INSA-Lyon, Villeurbanne; CNRS, UMR5240, Microbiologie Adaptation et Pathogénie, Lyon, France

OPEN ACCESS Freely available online

PLoS PATHOGENS

Structural and Functional Insights into the Pilotin-Secretin Complex of the Type II Secretion System

Shuang Gu[1☯], Saima Rehman[1☯], Xiaohui Wang[2], Vladimir E. Shevchik[2]*, Richard W. Pickersgill[1]*

1 Queen Mary University of London, School of Biological and Chemical Sciences, London, England, **2** Université de Lyon, Université Lyon 1, Lyon; INSA-Lyon, Villeurbanne; CNRS, UMR5240, Microbiologie Adaptation et Pathogénie, Lyon, France

Abstract

Gram-negative bacteria secrete virulence factors and assemble fibre structures on their cell surface using specialized secretion systems. Three of these, T2SS, T3SS and T4PS, are characterized by large outer membrane channels formed by proteins called secretins. Usually, a cognate lipoprotein pilot is essential for the assembly of the secretin in the outer membrane. The structures of the pilotins of the T3SS and T4PS have been described. However in the T2SS, the molecular mechanism of this process is poorly understood and its structural basis is unknown. Here we report the crystal structure of the pilotin of the T2SS that comprises an arrangement of four α-helices profoundly different from previously solved pilotins from the T3SS and T4P and known four α-helix bundles. The architecture can be described as the insertion of one α-helical hairpin into a second open α-helical hairpin with bent final helix. NMR, CD and fluorescence spectroscopy show that the pilotin binds tightly to 18 residues close to the C-terminus of the secretin. These residues, unstructured before binding to the pilotin, become helical on binding. Data collected from crystals of the complex suggests how the secretin peptide binds to the pilotin and further experiments confirm the importance of these C-terminal residues *in vivo*.

Citation: Gu S, Rehman S, Wang X, Shevchik VE, Pickersgill RW (2012) Structural and Functional Insights into the Pilotin-Secretin Complex of the Type II Secretion System. PLoS Pathog 8(2): e1002531. doi:10.1371/journal.ppat.1002531

Editor: Thomas C. Marlovits, IMP/IMBA Research Center, Austria

Received June 17, 2011; **Accepted** December 28, 2011; **Published** February 9, 2012

Funding: We acknowledge funding from BBSRC, HEFCE and Queen Mary University of London. V. S. was supported by a grant from French ANR-2010-BLANC-1531 SecPath programme. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: r.w.pickersgill@qmul.ac.uk (RWP); vladimir.shevchik@insa-lyon.fr (VES)

☯ These authors contributed equally to this work.

Introduction

The secretins are an important group of bacterial membrane proteins whose function is to facilitate the transport of secreted proteins and macromolecular complexes across the outer membrane [1].They are essential components of the type II and type III secretion systems (T2SS and T3SS respectively) and play a key role in the assembly of type IV pili (T4P) and release of filamentous bacteriophages. Determination of the structure of secretins has been confined to low-resolution transmission electron microscopy and cryo EM studies [2–5] which show membrane penetrating ring structures with 12–14 fold rotational symmetry. A specialized class of small lipoprotein pilotins bind their cognate secretins and facilitate oligomerization, insertion and proper assembly in the outer bacterial membrane. In this paper we explore the structure and function of the pilotin from *D. dadantii* (OutS), several other pilot proteins have been described [6–12]. Pilotins whose structures have been determined are MxiM (PDB code: 1Y9L) of the T3SS of *Shigella flexneri* and PilW/PilF (PDB codes: 2VQ2)/2FI7 and 2HO1) of the T4P of *Neisseria meningitidis* or *Pseudomonas aeruginosa* [13,14]. The cracked β-barrel structure of MxiM has been solved in complex with an 18 residue peptide from the cognate secretin MxiD (PDB code: 2JW1) and the authors propose a model for the way MxiM assists MxiD assembly [10,15]. The other known pilot structure, PilW/PilF, appears to perform a broadly similar function to MxiM, ensuring multimerization of the secretin PilQ into the outer membrane, but has a different architecture comprising six serial α-helical tetratricopeptide repeats [9,12]. A third auxiliary secretin-binding protein has been characterised structurally, PilP, which is also involved in the assembly or stability of the secretin PilQ of *Pseudomonas aeruginosa* [16], the structure comprises a sandwich of two sheets each with three anti-parallel β-strands.

The type II secretion system spans both the inner and outer bacterial membranes [17,18]. It consists of an inner membrane subcomplex, periplasmic pseudopilins and the outer membrane secretin [19]. There have been considerable recent advances in our understanding of the T2SS secretin. First the structure of the N-terminal periplasmic domains N0, N1, and N2 in complex with a nanobody [20] was elucidated (PDB code 3EZJ) and secondly a cryo EM reconstruction of the secretin itself has been described [3]. In the absence of pilotin the *D. dadantii* secretin (OutD) mislocates to the inner membrane [21]. The pilotin possesses at its N-terminus the characteristic lipoprotein signal sequence (LAAC), with the signal peptidase LspA cleaving site just before the cysteine to which the lipid is covalently attached. The pilotin (OutS) binds to the C-terminal 62 residues of the secretin (OutD) [11,19]. Here we elucidate the structure of the T2SS pilotin and show that it binds tightly to 18 residues close to the C-terminus of the secretin subunit causing this unstructured region to become helical on forming the complex.

Author Summary

Pathogenic bacteria deliver toxins and virulence proteins into host cells and tissues using specialised secretion systems such as the type II and type III secretion systems. These secretion systems have a pore formed by secretin protein subunits through which the disease causing protein effectors and toxins pass. The secretin must be targeted to and assembled in the outer-membrane and a pilotin protein facilitates this process. In the absence of the pilotin the secretin is degraded or mislocates to the inner-membrane, in either case the secretion system in non-functional and the bacterium cannot cause disease. Here we show how the pilotin and the secretin of the type II secretion system interact, these insights may be useful for the development of antibacterial compounds to interfere with secretin targeting and assembly and defeat pathogenic bacteria such as *Vibrio cholerae* and enterotoxigenic *E. coli* which infect man and *Dickeya dadantii* which threatens food security.

Results/Discussion

Structure of the T2ss Pilotin

To ensure authentic folding and production of soluble *Dickeya dadantii* pilotin (OutS) in the *E. coli* periplasm, the PelB secretion sequence was substituted in place of the N-terminal lipidation sequence thereby preventing lipidation of the pilotin. This substitution facilitated protein production and crystallization without compromising secretin-binding [11]. Cleavage of the secretion signal accompanies transport in to the periplasm. The crystal structure of the pilotin was determined using the anomalous scattering from a potassium tetrachloroplatinate derivative and the structure refined at 1.65 Å resolution (Table 1). The two copies of the pilotin subunit in the asymmetric unit of the crystal are virtually identical in structure (root mean square deviation of 94 α-carbon atoms, residues 38 to 132, is 0.243 Å). The structure is clearly defined in the electron density map except for the N-terminal residues preceding Val 38 which presumably form a flexible linker to the lipidation site. The architecture of the pilotin is the remarkable insertion of one α-helical hairpin into a second open α-helical hairpin with bent final helix (Figure 1); this is unlike the two other pilotin structures solved and is profoundly different from any previously described four helix bundle. The first helix of five turns (residues 40–60), is connected to the second of four turns (residues 69–82) by an 8 residue loop. The second loop of 10 residues connects to the third helix of four turns (residues 93–106) which packs against helix one. A short three residue loop which connects helices three and four and the disulfide bridge, between Cys 115 in the second turn of helix four and Cys 61 the first residue of the helix one to helix two loop (Figure 1B), sets the scene for the packing of helix four (residues 111–130). The pronounced bend of 65° in helix four is important for the architecture; the helix has three large hydrophobic residues which are at least partially buried by interactions with hydrophobic residues on the other three helices: Phe 118, Met 122, and Phe 125. The requirement to pack conserved Phe 125 appears to dictate the severe bend of this helical element. In the crystal the pilotin subunits form a dimer, with Arg 63 and Asn 119 (Figure 1B) involved in an electrostatic interface, between subunits, however there is currently no evidence that dimerization occurs in solution or *in vivo*.

Sequence Similarity and Structural Similarity

The majority of the 13 absolutely conserved residues in the sequence alignment (Figure S1A) appear to be of structural rather than of direct functional significance. The two highly conserved cysteine residues, 61 and 115, form the disulfide bridge between helices α1 and α4 that stabilizes the correct nested α-helical protein fold of the pilotin is functionally relevant. When a reducing agent was used in pull-down assays or in bacterial two-hybrid tests, the pilotin was unable to bind the cognate secretin (Figure S6). Presumably this is because the disulfide is reduced and the pilotin did not fold correctly in the cytoplasm. Interestingly, the previous mutagenesis analysis revealed several structurally or functionally relevant residues of pilotin (OutS), notably conserved Leu 57, Arg 63 and Ser 97 [11]. Substitution each of these prevented secretin (OutD) targeting to the outer membrane. Conserved Cys 21 covalently attaches to the lipid and this residue is essential for targeting the pilotin to the outer membrane. Conserved Gln 46, on the solvent exposed surface of helix 1 (Figure 1), must also be of functional rather than structural significance and maps to the extremity of the concave surface of the pilotin formed by helices α1, α3 and α4; it is plausible that this concave surface is the binding site for the secretin (Figure 1B). Conserved residues including: Gln 46 and Leu 104 and semi-conserved Leu/Val/Ile 50, Phe/Leu 118 are also in this region. Residues 49, 52 and 53 on the solvent exposed surface of helix α1 are Ser/Ala for the former two and Ala/Gly for the latter, respectively (Figure S1A). This conservation of small residues at these positions is consistent with this region being important in binding as there is no structural reason why larger residues could not be accommodated at these sites unless the secretin binds tight up against the first helix (Figure 1). A DaliLite database search revealed that P40 nucleoprotein has a similar arrangement of α-helices to that of the pilotin. The Dali score was 6.0 and sequence identity 5%. The P40 nucleoprotein domain architecture is however substantially more complex with seven helices instead of the pilotin's four. The concave surface of the corresponding helices of P40 nucleoprotein is occupied by a helix supporting the view that this is the binding site for an α-helix. Mutations of the pilotin binding surface confirm the importance for binding of some of the residues decorating the concave surface (Table S1). Mutating Ser 49 to Arg has a profound effect on binding and the mutants Leu 96 Ala, Leu 100 Ala and Gln 114 Ala have a substantial effect on binding as expected if this is the binding surface (Figure 1B).

Secretin-Binding to the Pilotin in Solution

It had previously been shown that the C-terminal 62 residues of the secretin bound to the pilotin [11]. This secretin peptide was produced with ^{15}N-label as a fusion to GST and then released with PreScision protease. The backbone amide protons were poorly dispersed in the ^{1}H-^{15}N HSQC spectra revealing the C-terminal 62 residues are unstructured in solution (Figure S4). NMR cross-titration studies revealed that only peaks corresponding to residues 691–708 of the secretin peptide were shifted on addition of unlabelled pilotin (Figure S2; with assignment of secretin peptide shown in Figure S3). When the pilotin was ^{15}N-labelled, good dispersion of the backbone amides protons was observed as expected given its folded structure (Figure 2). Titration of the unlabelled 62 residue secretin peptide into the ^{15}N-labelled pilotin produced a large number of peak shifts (Figure 2A). Shift perturbations are extremely sensitive indicators of structural changes and the extent of the changes observed is compatible with the secretin peptide decorating the surface of the pilotin and causing subtle structural rearrangements, perhaps in packing interactions in the hydrophobic core, reflected in chemical shift

Table 1. Crystallographic data and refinement statistics for pilotin.

Data Collection	**Native**	**Tetrachloroplatinate derivative**
Space group	$P2_12_12_1$	$P2_12_12_1$
Cell parameters (Å)	*a = 49.7, b = 53.1, c = 98.7*	*a = 49.8, b = 51.8, b = 97.9*
Molecules per asymmetric unit	2	2
Platinum sites/au	0	4
Wavelength (Å)	1.0718	1.0718
Resolution (Å)	46.78–1.65 (1.71–1.65)[a]	52.81–2.90 (3.06–2.90)[a]
Number of unique reflections	32440 (4652)[a]	6121 (873)[a]
Multiplicity	7.8 (6.8)[a]	12.6 (13.4)[a]
Completeness (%)	100.0 (99.9)[a]	99.9 (100.0)[a]
R_{merge} (%)[b]	0.200 (0.336)[a]	0.103 (0.214)[a]
Mean I/sigma (I)	6.9 (2.6)[a]	19.2 (12.3)[a]
R_{pim} (%)[c]	0.063 (0.140)[a]	0.033 (0.062)[a]
MSAN[d]	-	1.20
Wilson B-factor (Å^2)	21.4	60.1
Refinement		
Resolution limits (Å)	46.8–1.65	
Reflections (work/test)	30555/1578	
R-factor/R-free[e] (%)	0.197/0.249	
rmsd bond(Å)/angle (°)	0.006/0.918	
Number of protein (solvent) atoms	1465 (195)	
Average B-factor protein (solvent) (Å^2)	30.9 (46.0)	
Ramachandran plot statistics (%)		
Residues in most favoured regions	98.4%	
Residues in additional allowed regions	1.6%	

[a]The parameter values for the range 1.85–1.76 Å and 3.06–2.90 Å are given in parentheses for native and heavy metal derivative data, respectively.
[b]$R_{merge} = \Sigma_{hkl} \Sigma_i|I_i - <I>|/\_\Sigma_{hkl} \Sigma I_i$, where I_i is the intensity of the i^{th} observation, <I> is the mean intensity of the reflection, and the summations extend over all unique reflections (hkl) and all equivalents (i), respectively.
[c]$R_{pim} = \Sigma_{hkl} [n/(n-1)]^{1/2} \Sigma_i|I_i(hkl) - <I(hkl)>|/\Sigma_{hkl} \Sigma_i I_i(hkl)$, where n is the multiplicity, other variables as defined for R_{merge} [46].
[d]MSAN is the Mid slope of Anomalous normal probability.
[e]R-factor $= \Sigma_{hkl}|F_o - F_c|/\Sigma_{hkl} F_o$, where F_o and F_c represent the observed and calculated structure factors, respectively. The R-Factor is calculated using 95% of the data included in refinement and R-free the 5% excluded.
doi:10.1371/journal.ppat.1002531.t001

changes across much of the structure. The secretin peptide binds to the pilotin in a 1:1 stoichiometric ratio as determined from the NMR titration where the intensity of the shifted peaks of the complex saturate at an equimolar ratio of OutD to OutS. Since only residues 691–708 of the secretin were affected by interaction with pilotin (Figure S2), a synthetic 18 residue peptide corresponding in sequence to these residues was assessed. The pattern of shifts in ^1H-^{15}N HSQC pilotin spectrum observed using this synthetic 18 residue peptide was identical to that using the 62 residue secretin peptide (Figure 2B, C) confirming that it is these 18 residues that are those principally involved in binding to the pilotin. Circular dichroism measurements also showed the unstructured nature of the secretin peptide and provided evidence that the peptide becomes helical on binding (Figure 3A). The signal saturates at a stoichiometric ratio of secretin peptide to pilotin. The CD spectra of the pilotin and secretin peptide together correspond to more helical structure than the spectra of the pilotin and secretin peptide summed. The additional helical content can be quantified as 12 residues assuming all secretin and pilotin molecules are in complex, a reasonable assumption given the high affinity of complex formation (see below). The most plausible explanation for this is that 12 residues of the secretin peptide become helical on binding to the pilotin. The helical propensity of the 18 residue secretin peptide was apparent from secondary structure predictions (Jpred [22] and shown on Figure S1B). To estimate the binding affinity of pilotin for the secretin peptide, fluorescence spectroscopy was used. Since the 18 residue secretin peptide has no tryptophan residues, quenching of the fluorescence signal from the single tryptophan residue in the pilotin on addition of the secretin peptide, was used to determine the affinity of binding. The 1:1 stoichiometric binding ratio can be seen from the saturation of the fluorescence quenching of OutS by an equimolar quantity of the secretin peptide (Figure 3B). The binding of the 18 residue peptide is tight with K_d of 55 ± 20 nM (Figure 3B) and is comparable to that of the T3SS pilotin-secretin complex [10,15]. 3JHNHA spectra of the complex showed peaks coupled by less than 5 Hz (Figure S5) confirming that at least four residues of the secretin peptide become helical on binding to the pilotin, there may be more, but they are hidden by overlapping peaks.

Model of the Pilotin/Secretin Peptide Complex

Crystals of the pilotin/secretin complex were grown, they belong to space group $P6_5$ but are twinned (Table S2) and have solvent channels with disordered density within. Nevertheless four

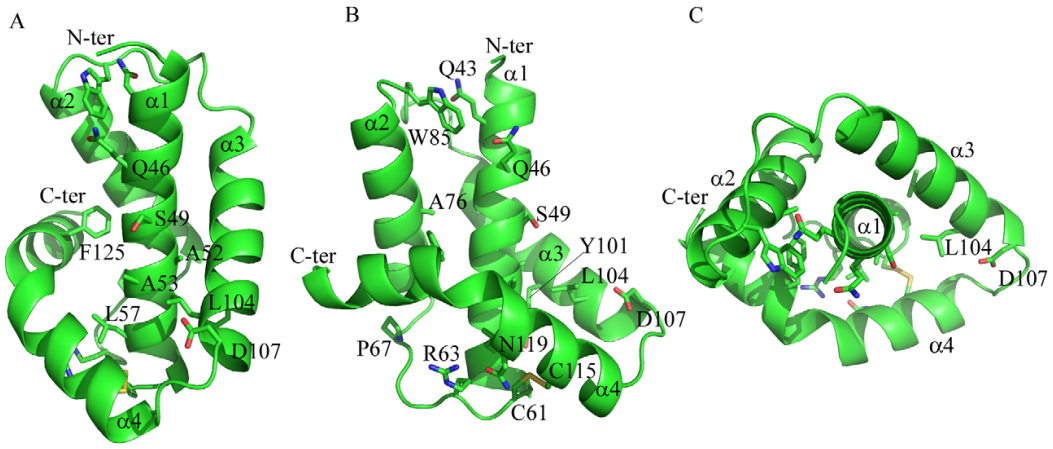


Figure 1. Structure of the pilotin. The crystal structure of *Dickeya dadantii* OutS consists of four α-helices, the last of which is bent. (A) α3 and bent α4 wrap around the anti-parallel hairpin formed by α1 and α2. Conserved residues (see Figure S1 for sequence alignment) are represented as sticks. (B) Rotated about the y-axis by 90°, this view reveals the concave surface formed between helix α1 and helices α3and α4. The disulfide can be seen linking α1 and α4. (C) Rotated around the x-axis by 90°, α1 is surrounded by the other three helices. A52, A53, S49 and D107 are not strictly conserved but the first three are always small and 107 is either D or E suggesting some functional constraint on this region. Q46 is absolutely conserved and may map the other extremity of the binding site. Figure 1 and panels D and E of Figure 3 were produced using PYMOL.
doi:10.1371/journal.ppat.1002531.g001

pilotin subunits can be found by molecular replacement and density for peptide can be seen occupying the concave surface of the pilotin. The evidence that suggests residues 694 to 704 of the secretin adopt a helical conformation in the electron density map, a part of the simulated annealed omit map is shown in Figure 3C. Ten residues are in helical conformation and the hydrophobic surface of the amphipatic helix interacts with the hydrophobic surface of the pilotin (Figure 3D, E). In this model the methyl groups of T692 and V697 interact with L100 and L104, I701 interacts with L50 and M122; F704 (with L47, V79 and F125) and F694 (with F118, L104 and Q114) interactions occur either side of these central interactions. D107 is the N-terminal helix capping residue, stabilizing the helix dipole of the bound secretin peptide. The peptide binds tight up against Q46, S49, A52 and A53, providing an explanation for their conservation or presence only as small residues in the case of the latter three. Interestingly, the same dimer as described for the non-complexed secretin is seen in these crystals too suggesting that the interaction may have some biological relevance. The quality of the refinement is relatively poor because of the disordered protein and for that reason the structure is being referred to as a model of the secretin peptide/pilotin interaction.

Intrinsically disordered regions of proteins such as these C-terminal residues of the secretin subunit facilitate binding by increasing their capture radius for cognate partner, the so-called fly-casting mechanism [23]. Initial weak binding may draw the secretin and pilotin together and as the secretin peptide folds on the pilotin surface the binding becomes tighter, locking the two together.

Assessment of the Pilotin/Secretin Interaction In Vivo

A series of *in vivo* experiments were used to test the proposed model of secretin binding to the pilotin. The 62 residue C-terminus of secretin possesses three putative α-helices (Figure 4A) [22]. If the region consisting of the first two C-terminal helices were deleted, the truncated OutDΔC1 behaves like wild type secretin. Firstly, OutDΔC1 was barely detectable in the absence but was well produced in the presence of pilotin OutS (Figure 4B). Secondly, with pilotin, the mutant secretin was mainly recovered in the outer membrane fractions (Figure 4C). Consistent with the outer membrane location, expression of OutDΔC1 in the presence of OutS results in a rather low level of *pspA* induction. Phage shock protein (psp) response helps to maintain proton motive force in cells under pmf-dissipating stress and is indicative of mislocalization of the secretins in the inner membrane [8,11]. These results are consistent with the NMR experiments demonstrating that it is the 18 residue C-terminus of OutD which binds tightly to the pilotin. Despite its outer membrane location, OutDΔC1 was unable to restore pectinase secretion in *D. dadantii* Δ*outD* A3559 strain (data not shown) indicating an important functional relevance of the deleted region. Deletion of the extreme C-terminus of secretin resulted in partial stabilization of the truncated OutDΔC2 secretin as judged from the quantity produced in the absence of the pilotin, but prevented its correct targeting to the outer membrane (Figure 4B,C). In the presence of pilotin, the amount of OutDΔC2 was increased indicating that the pilotin can still stabilize and hence interact with the truncated secretin but is not able to target it to the outer membrane. In agreement with this, expression of OutDΔC2 strongly induced *pspA* even in the presence of pilotin. Deletion of the full 62 residue C-terminus of OutD resulted in neither stabilization nor the correct targeting of the truncated secretin OutDΔC3 as judged by low protein content but high *pspA* level (Figure 4B).

Conclusion

The type II secretin/pilotin complex from *Klebsiella oxytoca* has been imaged by cryo electron microscopy at modest resolution [24]. The secretin subunits form a dodecameric ring with relatively weak radial arms that the authors tentatively assign as the pilotin or pilotin bound to the secretin C-terminus [4,24]. Comparing the envelope of the complex with that determined by Reichow et al. (2010) of the secretin only [3], the radial arms are located either in the periplasm or in the inner leaflet of the outer membrane. This is the position anticipated from the observation in this work that the

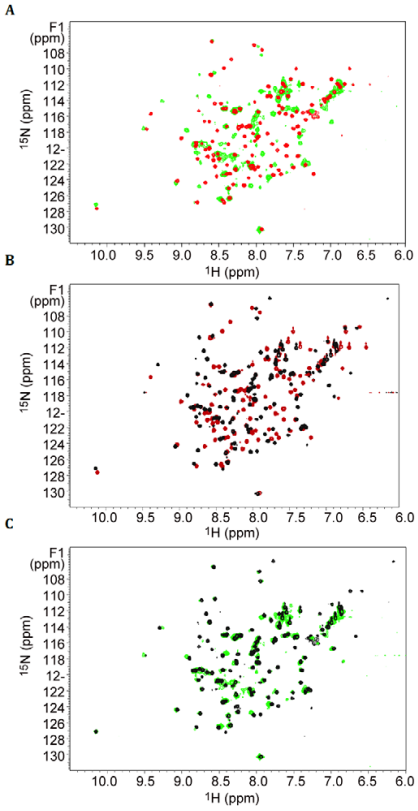


Figure 2. Elucidation of secretin-pilotin interactions. Titration of secretin peptides into ^{15}N labelled pilotin. ^{1}H-^{15}N-HSQC spectra of the pilotin in the absence of secretin (red spectra), in the presence of 62 residue secretin peptide (green) and in the presence of 18 residue peptide (black). Protein concentration was 100 µM. (A) Pilotin in the absence and presence of 62 residue secretin peptide. (B) Pilotin in the absence and presence of the 18 residue secretin peptide. (C) Overlay of the two complexes with secretin peptides showing the 18 residue peptide is behaving in a closely similar way to the 62 residue peptide. doi:10.1371/journal.ppat.1002531.g002

pilotin interacts tightly with 18 residues close to the C-terminus of the secretin subunit, so it is entirely plausible that the radial arms seen in the cryo-EM map correspond to the unusual four helix bundle of the type II secretion pilotin bound to the induced C-terminal helix of the secretin subunit.

Recent structural studies have revealed striking structural similarities between components from distant secretion systems and other bacterial cell machines. Besides the expected structural homologies between several conserved components of ancestrally related T2SS and T4P [25–27] several other remarkable structural similarities should be mentioned. Notably, the extreme N-terminal N0 domain is shared by secretins from the T2SS, T3SS and T4P but is also structurally related to a domain of lipoprotein DotD from the T4SS, a domain of VgrG from the T6SS and from a TonB dependent receptor FpvA [20,28–30]. Similarly the N1/N2 domains of secretins show a significant structural homology with several ring-forming proteins from the T3SS [20,31,32]. It is therefore becoming common to attribute similar function to the proteins or domains of a related structure. Given this background it is remarkable then that the T2SS pilotin described here is profoundly different in architecture to the T3SS pilotin but has similar function. Both the T2SS and the T3SS pilotin bind the extreme C-terminal region of their cognate secretins and this previously unstructured part of the secretin becomes an ordered α-helix on binding. It is therefore remarkable that the corresponding pilotins are different in architecture, one an open β-barrel (T3SS; MxiM) the other an unusual helical bundle (T2SS; OutS). These striking structural differences show that in these systems the pilotins have been evolved independently to play similar roles.

Materials and Methods

Plasmid Construction, Expression, Purification and Protein Analysis

The pET-20b(+) plasmid expressing non-lipidated OutS (residues 26 to 132) fused to N-terminal PelB signal peptide has been constructed previously [11] at a manner as the sequence of mature non-lipidated OutS after cleavage by the signal peptidase LepB is: *MDP*^{26}VKNT etc. To fuse a C-terminal 6His tag to non-lipidated OutS, *Sal*I site was introduced at the end of *outS* sequence by using the primer (5′-ctt gac gcc atg cgc acc gt**c ga**c tga ggg gga agc aac tgc) and the reverse complementary one (mutated bases are underlined). Then, by *Sal*I/*Xho*I digestion the sequence coding for non-lipidated OutS was fused with that coding for 6His. Mutants of OutS were made using Strategene QuikChange and confirmed by sequencing.

To generate OutD truncated derivatives, an *Eco*47III site and V678A substitution were introduced using the primer (5′-gcgcggcgaaggcaacggagcgctggataacaacaccctgc) and the reverse complementary one. This site and naturally existing *Nru*I and *Psi*I sites were used to generate OutDΔC1 (Δ650–678), OutDΔC2 (Δ679–705) and OutDΔC3 (Δ650–705) derivatives. To fuse the C-terminal segments of OutD to GST, the corresponding gene fragments were subcloned from pTdB-OD plasmids expressing either OutD, or OutDΔC1, or OutDΔC2, or OutDΔC3 into pGEX-6P-3 or pGEX-3X vectors in frame with the GST coding sequence.

Protein Expression, Purification and Analysis

E. coli BL21(DE3) strain (Stratagene) was used to produce non-lipidated pilotin (OutS) and GST-secretin (OutD) derivatives. Non-lipidated OutS was released from the periplasm by osmotic shock as described previously [33] and purified by size-exclusion chromatography Superdex S75 10/300 GL (GE Healthcare). The OutD peptide was purified and then released from GST-OutD fusion as described previously [34]. For NMR spectroscopy, uniformly ^{15}N- and ^{13}C-labeled pilotin and secretin peptides were produced by growing cell cultures in M9 minimal medium that contained ^{15}N-ammonium chloride *and* ^{13}C-*D-glucose* (Cambridge Isotope Laboratories Inc.) as the sources of nitrogen and carbon, respectively. The 18 residue synthetic secretin peptide (residues 691 to 708 inclusive) was purchased from Generon. Cell membrane fractionation by sucrose density gradient centrifugation was performed as described previously [11] with steady-state cultures of *E. coli* NM522 (Stratagene) expressing OutD derivatives from pTdB-OD and OutS from pACT-S plasmid. The location of outer membrane porins was determined by staining with Coomassie G-250. The position of inner membrane fractions was estimated by immunoblotting with TolA antibodies and NADH oxidase activity. *E. coli* MC3 strain carrying a *pspA-lacZ* fusion [35] was used to estimate miss location of OutD derivatives. To assess functionality of OutD derivatives, complementation assays with *D. dadantii* Δ*outD* A3559 strain were used as previously described [36]. SDS-PAGE and immunoblotting were performed

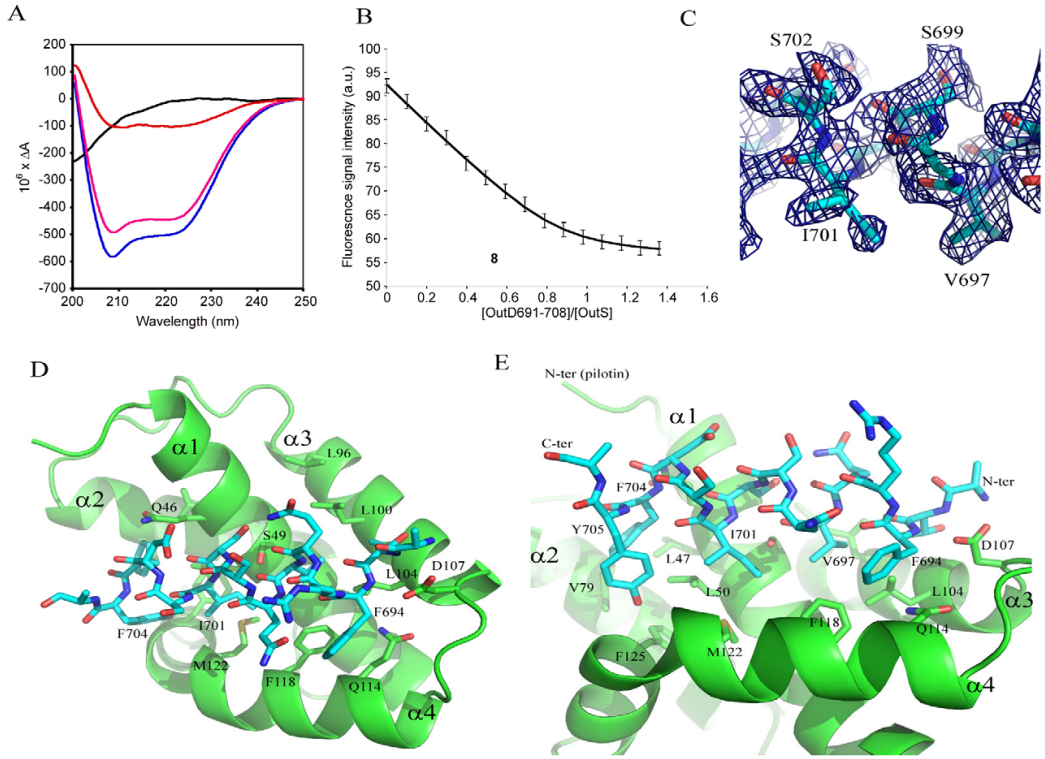


Figure 3. The secretin-pilotin complex. (A) The far uv circular dichroism spectra of the 18 residue secretin peptide alone (black), pilotin alone (pink) and a stoichiometric ratio of both pilotin (OutS) and secretin peptide ($OutD^{691-708}$) together (blue). The difference between the secretin/pilotin complex and the pilotin only is shown in red. As forming the stoichiometric ratio diluted both proteins by half, these data were multiplied by two to compensate for the dilution factor. The concentration of both pilotin and secretin peptide were 0.55 mM. 3JHNHA evidence of helical conformation is presented in Figure S5. (B) Measurement of the binding affinity of the pilotin OutS for the 18 residue secretin peptide determined using fluorescence spectroscopy. 1 μM pilotin was titrated with 50 μM secretin peptide in to 1 μM pilotin so that there was no dilution of pilotin. The stoichiometry can be seen to be 1:1 within experimental error. The K_d is 55±20 nM. Details of the equation fitted can be found in Table S1. (C) Part of the simulated annealed omit map showing the quality of the electron density map used to derive the models shown in (D) and (E). (D) Model of the secretin peptide bound to the pilotin, D107 acts as an N-terminal helical cap. (E) Close up showing the hydrophobic nature of the complementary side chains involved in forming the complex.
doi:10.1371/journal.ppat.1002531.g003

as previously [19]. Anti-OutD rabbit serum was raised against entire OutD purified from a recombinant *E. coli* strain. Anti-TolA serum was kindly provided by J.C. Lazzaroni.

Crystallisation and Structure Determination

Hampton Research sparse matrix screen was used to search for crystallization conditions. Crystals were grown using hanging drop vapour equilibration using 10 mg/ml OutS and a reservoir of 2 M ammonium sulphate, 2% PEG 400 and 0.1 M HEPES pH 7.5. Vitrification of crystals in liquid nitrogen was achieved using the reservoir solution with 2.1 M ammonium sulphate and augmented with 15–25% glycerol. Data were collected at ESRF ID23-1 and processed using MOSFLM [37] and scaled using SCALA [38]. SAD data were collected from a crystal soaked in 25 mM potassium tetrachloroplatinate (K_2PtCl_4) for 4 days. The structure was solved using PHENIX [39] and COOT [40] and refined using the native data at 1.65 Å resolution and non-crystallographic symmetry restraints. The final model comprises 188 amino acid residues and 208 water molecules. DALILITE [41] was used to search for similar structures, CLUSTALW [42] for sequence alignment and JPRED for secondary structure predictions [22]. Crystals of the complex were grown using a 1.0:1.1 molar ratio of pilotin: secretin peptide and crystallized using a reservoir of 2 M ammonium sulphate, 0.1 M Tris, pH 8.5. Around 100 complex crystals were screened before a well-diffracting reasonable ordered crystal was found. The pilotin/secretin complex was solved by molecular replacement using data collected at DLS I02 and CCP4/PHENIX/COOT. These crystals appear to be $P6_522$ but are most likely twinned $P6_5$ with four pilotin molecules in the asymmetric unit. The packing of the molecules is such that there are large solvent channels running through the crystal lattice. These solvent channels appear to have disordered protein present the modelling of which hampers refinement. The disordered regions do not gain clarity if the lower symmetry space group $P3_2$ is used (for more details see Table S2).

Preparation of Proteins for Nmr Spectroscopy and Acquisition of Nmr Spectra

Samples of 0.05 to 0.5 mM labelled proteins in 90% H_2O, 10% ^2H_2O containing 20 mM Tris (pH 7.0) and 150 mM NaCl. All

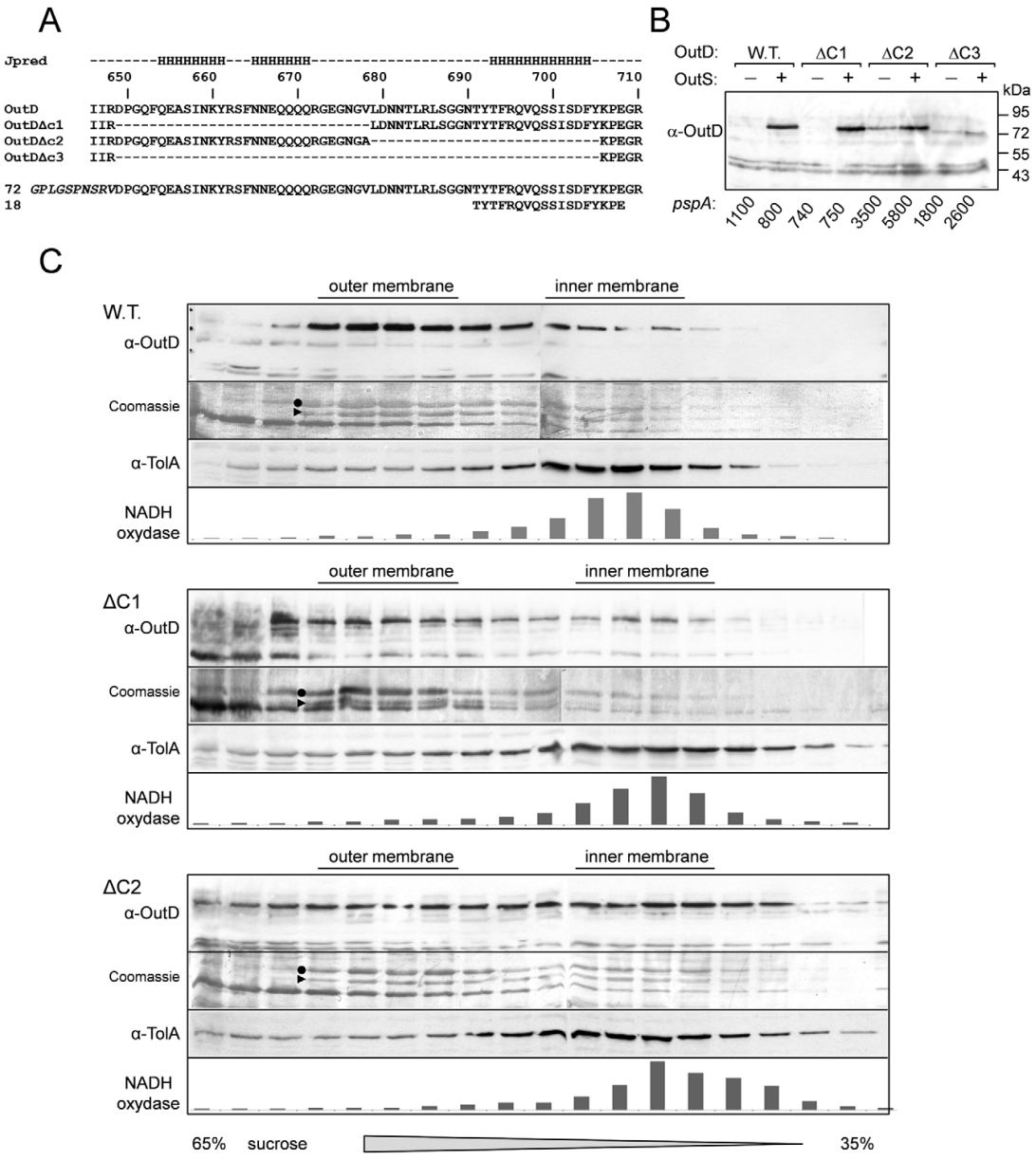


Figure 4. ***In vivo*** **determination of secretin interactions with pilotin.** Removal of the extreme C-terminal region of the secretin results in mislocalization of the secretin to the inner membrane (A) Sequence of the C-terminus of *Dickeya dadantii* secretin (OutD) along with the secondary structure prediction for this region; the predicted three helices are marked "H". The sequences of the three secretin deletion mutants (OutDΔC1 to OutDΔC3) and the C-terminal secretin peptides used are shown. (B) Stabilization of truncated derivatives of the secretin (OutD) by pilotin (OutS) *in vivo*. *E. coli* MC3 cells expressing an OutD derivative (indicated above) and either OutS (+) or empty pACT3 vector (−) were grown for 12 h at 30°C in LB medium and then analyzed by immunoblotting with OutD antibodies. In the same cultures, β-galactosidase activity was assessed to estimate expression of *pspA-lacZ*. An elevated level of *pspA* reflects mislocalization of the corresponding secretin derivative to the inner membrane. Equivalent amounts of cell extracts were loaded into each well and used for activity measurement. (C) Pilotin promotes the outer membrane location of the full-length secretin OutD and truncated secretin OutDΔC1 but not OutDΔC2. The whole membrane fraction from *E. coli* NM522 cells coexpressing the indicated secretin derivatives and pilotin was separated by flotation sucrose gradient centrifugation and analyzed by immunoblotting with OutD-antibodies or stained with Coomassie G-250 to detect the major porins, which reflect the position of the outer membrane. Immunoblotting with TolA-antibodies and NADH-oxydase activity indicate the position of the inner membrane fractions. OmpA is indicated by a triangle and OmpC/F by a dot.

doi:10.1371/journal.ppat.1002531.g004

NMR spectra were acquired at 15°C using Bruker Avance 700- and 600-MHz spectrometers. Assignment of ^{1}H, ^{15}N, and ^{13}C resonances of the backbone was achieved by analysis of HNCACB, CBCA (CO)NH triple resonance experiments [43].

Circular Dichroism

Far-UV CD measurements were made using a Jasco J-715 spectropolarimeter equipped with a PTC-348WI temperature controller. Spectra were recorded in 20 mM Tris, 150 mM NaCl (pH 7.0) at 15°C using 1 mm path length fused silica cuvettes. The spectra are presented as differential absorbance after baseline subtraction. Calculations employed CONTIN, SELCON, and CDSSTR [44].

Fluorescence Spectroscopy

Fluorescence data were collected using a Jasco FP-6300 Spectrofluorometer. To avoid exciting tyrosyl side chains, an excitation wavelength of 290 nm was used. Emission spectra were recorded at 15°C in steps of 2 nm from 310 to 400 nm. The fluorescence signal at 340 nm was plotted to calculate K_d. Pilotin spectra was measured at 1 μM. 50 μM of secretin peptide prepared in 1 μM pilotin was titrated into pilotin solution (for more information see also Table S1 and reference [45]).

Accession codes. Coordinates and structure factor amplitudes have been deposited in the protein databank with the accession codes 3UTK and 3UYM. The sequences of OutS and OutD are available in the UniProt database with accession codes Q01567 and Q01565, respectively.

Supporting Information

Figure S1 Sequence alignment of pilotins and the C-terminal region of their cognate secretins. The alignment of the pilotins is shown in panel (A) and alignment of the C-terminal region of their cognate secretins in panel (B). The position of the α-helices is indicated by H in the secondary structure row (predicted by Jpred for GspD). Shown are the OutS and OutD homologs of *Dickeya dadantii* (*Erwinia chrysanthemi* 3937), Q01567 and Q01565; *Pectobacterium carotovorum*, C6DAR0 and C6DAQ5; *Escherichia coli* O157:H7, Q7BSV3 and Q9ZGU0; *Klebsiella oxytoca*, P20440 and P15644; *Yersinia mollaretii*, C4S9G3 and C4S9F5; *Serratia odorifera*, D4E1I4 and A8GJQ5. Identical residues are in red, residues similar in character are green. Conserved residues are mapped on to the OutS pilotin structure in Figure 1 of the main text.
(DOC)

Figure S2 Spectroscopic analyses of secretin binding to the pilotin. 2D ^{1}H-^{15}N HSQC of ^{15}N labelled secretin peptide (OutD residues 649–685 and residues 649–710 for the major proteolytic fragment and minor full-length peptide, respectively) in the absence (black) and presence of pilotin (red). The concentration of secretin and pilotin were 50 μM and 100 μM, respectively. Both spectra were acquired using a Bruker 700 MHz at 15°C in buffer comprising 20 mM Tris pH 7.0, 150 mM NaCl and 10% $^{2}H_2O$.
(DOC)

Figure S3 Assignment of the backbone amide protons for the C-terminal secretin peptide. The data were acquired at 15°C using peptide in 20 mM Tris at pH 7.0, 150 mM NaCl and 10% $^{2}H_2O$ and a Bruker 700 MHz.
(DOC)

Figure S4 The C-terminus of the 62 residue secretin peptide ($OutD^{648–710}$) is unstructured. ^{1}H-^{15}N HSQC spectra of recombinantly produced ^{15}N-labelled secretin peptide (70 μM peptide in 20 mM Tris pH 7.0 with 150 mM NaCl at 15°C) acquired using a Bruker 700 MHz spectrophotometer. The low dispersion of the main chain amides reveals the peptide is intrinsically unstructured. The spectra are cleaner than those shown previously (Figure S2 and S3) because the spectra were acquired quickly. During more lengthy experiments the 62 residue secretin peptide is slowly cleaved degrading the quality of the spectra.
(DOC)

Figure S5 3JHNHA spectra of ^{15}N-labelled secretin peptide ($OutD^{680–710}$) in the absence (red) and presence (black) of pilotin (OutS). The spectra were acquired using a Bruker 700 MHz spectrophotometer at measured at 15°C. The new peaks, arrowed, have HN-HA coupling constants less than 5 Hz revealing that these residues are helical when bound. The 3JHNHA coupling constant was calculated according to measurement of the diagonal-peak to cross-peak intensity ratio in a 3D ^{15}N separated quantitative J-correlation spectra. These spectra show that at least four residues of the secretin peptide become helical on binding to the pilot.
(DOC)

Figure S6 Reducing environment prevents interaction of OutS with the C-terminal peptide of OutD. (A) GST pull-down assay shows that non-lipidated OutS does not bind to the GST-$OutD_{649–710}$ in reducing conditions. Soluble cell extracts of *E. coli* BL21(DE3) producing either GST alone (lane 1) or GST-$OutD_{649–710}$ (lanes 2 and 3) were combined with a periplasmic extract containing non-lipidated OutS, then loaded on Glutathione Agarose for 1 h and washed. The incubations were performed in either TBS (lanes 1 and 2) or TBS with 5 mM DTT (lane 3). Bound proteins were eluted with Laemmli sample buffer, separated by Tricine-SDS-PAGE and either stained (upper panel) or probed with OutS antibodies (lower panel). Asterisk indicates a degradation product of GST-$OutD_{649–710}$. (B) Bacterial two-hybrid assay (Karimova *et al.*, 1998) shows that OutS does not interact with GST-$OutD_{649–710}$ in the reducing conditions of the bacterial cytoplasm. The region coding for mature OutS (residues 26 to 137) was fused to the C-terminus of T18 subunit of Cya (pUT18-OutS) and the region coding for GST-$OutD_{649–710}$ was fused to the C-terminus of T25 subunit of Cya (pKT25-GST-Dct). When pUT18-OutS and pKT25-GST-Dct were coexpressed in *E. coli* DHP1 *cya* strain, the corresponding fusion proteins were well produced as shown by immunoblotting with anti-OutS and anti-GST antibodies, respectively. However, once plated on MacConkey-maltose agar, these bacteria generated white colonies (as did the empty vectors) and not red colonies (as produced by the known interacting OutC/OutC couple which was used as a positive control). T18-OutC is indicated by a dot and T25-GST-Dct by a triangle.
(DOC)

Table S1 Dissociation constants of secretin peptide from pilotin and pilotin mutants determined using fluorescence spectroscopy.
(DOC)

Table S2 Crystallographic data and refinement statistics for pilotin/secretin peptide complex.
(DOC)

Acknowledgments

X-ray data collection was at the European Synchrotron Radiation Facility, Grenoble, France using beamline ID23-1 and Diamond Light Source Oxfordshire using beamline I02. We are grateful to Drs Geoff Kelly and Steve Martin (National Institute of Medical Research, UK) for assistance with nuclear magnetic resonance and circular dichroism spectroscopy, respectively.

Author Contributions

Conceived and designed the experiments: RWP VES. Performed the experiments: SG SR XW VES RWP. Analyzed the data: SG SR XW VES RWP. Contributed reagents/materials/analysis tools: SG SR XW VES RWP. Wrote the paper: SG VES RWP.

References

1. Bayan N, Guilvout I, Pugsley AP (2006) Secretins take shape. Mol Micro 60: 1–4.
2. Collins RF, Frye SA, Kitmitto A, Ford RC, Tonjum T, et al. (2004) Structure of the Neisseria meningitidis outer membrane PilQ secretin complex at 12 angstrom resolution. J Biol Chem 279: 39750–39756.
3. Reichow SL, Korotkov KV, Hol WGJ, Gonen T (2010) Structure of the cholera toxin secretion channel in its closed state. Nature Struct Mol Biol 17: 1226–1233.
4. Chami M, Guilvout I, Gregorini M, Remigy HW, Muller SA, et al. (2005) Structural insights into the secretin PulD and its trypsin-resistant core. J Biol Chem 280: 37732–37741.
5. Hodgkinson JL, Horsley A, Stabat D, Simon M, Johnson S, et al. (2009) Three-dimensional reconstruction of the Shigella T3SS transmembrane regions reveals 12-fold symmetry and novel features throughout. Nat Struct Mol Biol 16: 477–485.
6. Burghout P, Beckers F, de Wit E, van Boxtel R, Cornelis GR, et al. (2004) Role of the pilot protein YscW in the biogenesis of the YscC secretin in Yersinia enterocolitica. J Bacteriol 186: 5366–5375.
7. Crago AM, Koronakis V (1998) Salmonella InvG forms a ring-like multimer that requires the InvH lipoprotein for outer membrane localization. Mol Micro 30: 47–56.
8. Hardie KR, Seydel A, Guilvout I, Pugsley AP (1996) The secretin-specific, chaperone-like protein of the general secretory pathway: Separation of proteolytic protection and piloting functions. Mol Micro 22: 967–976.
9. Koo J, Tammam S, Ku SY, Sampaleanu LM, Burrows LL, et al. (2008) PilF Is an Outer Membrane Lipoprotein Required for Multimerization and Localization of the *Pseudomonas aeruginosa* Type IV Pilus Secretin. J Bacteriol 190: 6961–6969.
10. Lario PI, Pfuetzner RA, Frey EA, Creagh L, Haynes C, et al. (2005) Structure and biochemical analysis of a secretin pilot protein. EMBO J 24: 1111–1121.
11. Shevchik VE, Condemine C (1998) Functional characterization of the *Erwinia chrysanthemi* OutS protein, an element of a type II secretion system. Microbiology 144: 1443219–3228.
12. Trindade MB, Job V, Contreras-Martel C, Pelicic V, Dessen A (2008) Structure of a widely conserved type IV pilus biogenesis factor that affects the stability of secretin multimers. J Mol Biol 378: 1031–1039.
13. Derrick J (2008) A Pilot Sheds Light on Secretin Assembly. Structure 16: 1441–1442.
14. Korotkov KV, Gonen T, Hol WGJ (2011) Secretins: dynamic channels for protein transport across membranes. Trends Biochem Sci 36: 433–443.
15. Okon M, Moraes TF, Lario PI, Creagh AL, Haynes CA, et al. (2008) Structural Characterization of the Type-III Pilot-Secretin Complex from *Shigella flexneri*. Structure 16: 1544–1554.
16. Golovanov AP, Balasingham S, Tzitzilonis C, Goult BT, Lian LY, et al. (2006) The solution structure of a domain from the *Neisseria meningitidis* lipoprotein PilP reveals a new beta-sandwich fold. J Mol Biol 364: 186–195.
17. Sandkvist M (2001) Biology of type II secretion. Mol Microbiol 40: 271–283.
18. Johnson TL, Abendroth J, Hol WGJ, Sandkvist M (2006) Type II secretion: from structure to function. FEMS Micro Letts 255: 175–186.
19. Shevchik VE, RobertBaudouy J, Condemine G (1997) Specific interaction between OutD, an Erwinia chrysanthemi outer membrane protein of the general secretory pathway, and secreted proteins. EMBO J 16: 3007–3016.
20. Korotkov KV, Pardon E, Steyaert J, Hol WGJ (2009) Crystal Structure of the N-Terminal Domain of the Secretin GspD from ETEC Determined with the Assistance of a Nanobody. Structure 17: 255–265.
21. Guilvout I, Chami M, Engel A, Pugsley AP, Bayan N (2006) Bacterial outer membrane secretin PulD assembles and inserts into the inner membrane in the absence of its pilotin. EMBO J 25: 5241–5249.
22. Cole C, Barber JD, Barton GJ (2008) The Jpred 3 secondary structure prediction server. Nucl Acid Res 36: W197–W201.
23. Shoemaker BA, Portman JJ, Wolynes PG (2000) Speeding molecular recognition by using the folding funnel: The fly-casting mechanism. Proc Natl Acad Sci U S A 97: 8868–8873.
24. Nouwen N, Ranson N, Saibil H, Wolpensinger B, Engel A, et al. (1999) Secretin PulD: Association with pilot PulS, structure, and ion-conducting channel formation. Proc Natl Acad Sci U S A 96: 8173–8177.
25. Sampaleanu LM, Bonanno JB, Ayers M, Koo J, Tammam S, et al. (2009) Periplasmic Domains of Pseudomonas aeruginosa PilN and PilO Form a Stable Heterodimeric Complex. J Mol Biol 394: 143–159.
26. Abendroth J, Rice AE, McLuskey K, Bagdasarian M, Hol WGJ (2004) The crystal structure of the periplasmic domain of the type II secretion system protein EpsM from Vibrio cholerae: The simplest version of the ferredoxin fold. J Mol Biol 338: 585–596.
27. Abendroth J, Kreger AC, Hol WGJ (2009) The dimer formed by the periplasmic domain of EpsL from the Type 2 Secretion System of Vibrio parahaemolyticus. J Str Biol 168: 313–322.
28. Nakano N, Kubori T, Kinoshita M, Imada K, Nagai H (2010) Crystal Structure of Legionella DotD: Insights into the Relationship between Type IVB and Type II/III Secretion Systems. PLOS Pathog 6: e1001129.
29. Leiman PG, Basler M, Ramagopal UA, Bonanno JB, Sauder JM, et al. (2009) Type VI secretion apparatus and phage tail-associated protein complexes share a common evolutionary origin. Proc Natl Acad Sci U S A 106: 4154–4159.
30. Garcia-Herrero A, Vogel HJ (2005) Nuclear magnetic resonance solution structure of the periplasmic signalling domain of the TonB-dependent outer membrane transporter FecA from Escherichia coli. Mol Micro 58: 1226–1237.
31. Yip CK, Kimbrough TG, Felise HB, Vuckovic M, Thomas NA, et al. (2005) Structural characterization of the molecular platform for type III secretion system assembly. Nature 435: 702–707.
32. Worrall LJ, Vuckovic M, Strynadka NCJ (2010) Crystal structure of the C-terminal domain of the Salmonella type III secretion system export apparatus protein InvA. Protein Sci 19: 1091–1096.
33. Fries M, Ihrig J, Brocklehurst K, Shevchik VE, Pickersgill RW (2007) Molecular basis of the activity of the phytopathogen pectin methylesterase. EMBO J 26: 3879–3887.
34. Login FH, Shevchik VE (2006) The single transmembrane segment drives self-assembly of OutC and the formation of a functional type II secretion system in *Erwinia chrysanthemi*. J Biol Chem 281: 33152–33162.
35. Bergler H, Abraham D, Aschauer H, Turnowsky F (1994) Inhibition of lipid biosynthesis induces the expression of the pspa gene. Microbiology 140: 1937–1944.
36. Bouley J, Condemine G, Shevchik VE (2001) The PDZ domain of OutC and the N-terminal region of OutD determine the secretion specificity of the type II out pathway of *Erwinia chrysanthemi*. J Mol Biol 308: 205–219.
37. Leslie AGW (2006) The integration of macromolecular diffraction data. Acta Cryst D62: 48–57.
38. Evans P (2006) Scaling and assessment of data quality. Acta Cryst D62: 72–82.
39. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, et al. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Cryst D66: 213–221.
40. Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. Acta Cryst D66: 486–501.
41. Holm L, Kaariainen S, Rosenstrom P, Schenkel A (2008) Searching protein structure databases with DaliLite v.3. Bioinformatics 24: 2780–2781.
42. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, et al. (2007) Clustal W and clustal X version 2.0. Bioinformatics 23: 2947–2948.
43. Bax A, Grishaev A (2005) Weak alignment NMR: a hawk-eyed view of biomolecular structure. Curr Opin Struct Biol 15: 563–570.
44. Sreerama N, Venyaminov SY, Woody RW (2000) Estimation of protein secondary structure from circular dichroism spectra: Inclusion of denatured proteins with native proteins in the analysis. Anal Biochem 287: 243–251.
45. Martin SR, Schilstra MJ (2008) Circular dichroism and its application to the study of biomolecules. In: Biophysical Tools for Biologists: Vol 1 in Vitro Techniques. San Diego: Elsevier Academic Press Inc. pp 263–293.
46. Weiss MS (2001) Global indicators of X-ray data quality. J Appl Cryst 34: 130–135.